Prelims: Introductory Calculus

2012 Michaelmas Term. Lecturer: Z. Qian
Mathematical Institute, Oxford

November 22, 2012

1 Standard integrals, integration by parts

It is important to grasp some basic techniques for evaluating integrals such as the method of substitutions, integration by parts etc. You should refer to Richard Earl's lecture notes, posted on the course web page, about a few standard substitutions.

Before we give some examples, let us introduce some notations. Recall that in A-level math, we write integral of a function f as $\int f(x)dx$ (indefinite integral), $\int_0^1 f(x)dx$ etc. $f(x)$ is called the integrand, and the expression $f(x)dx$ is called a *differential form* (of first order). It will be beneficial to introduce the notion of *differentials*. If $y = f(x)$ is a function of one variable x on some interval, then $dy = df(x) \equiv f'(x)dx$ is called the *differential* of f. The fundamental theorem in calculus says that

$$\int df(x) = \int f'(x)dx = f(x) + C$$

where C is an arbitrary constant.

The chain rule for derivatives implies that the differential of a function is *invariant* under substitutions. More precisely, suppose $y = f(x)$ is a function of x, making substitution $x = g(t)$ so that $y = f(g(t))$ is a function of t. Then $dx = g'(t)dt$ so that

$$\begin{aligned} dy &= f'(x)dx = f'(g(t))g'(t)dt \\ &= \frac{d}{dt}f(g(t))dt. \quad \text{[Chain rule]} \end{aligned}$$

That is $df(x) = df(g(t))$ if $x = g(t)$, in other words, when we work out the differential $df(x)$ it doesn't matter if we consider x as a variable or as a function of another variable. This principle also applies to differential forms of first order. The substitution method then can be summarized as the following equality

$$\begin{aligned} \int f(x)dx &= [\text{Substitute } x = \varphi(t)] \int f(\varphi(t))d\varphi(t) \\ &= \int f(\varphi(t))\varphi'(t)dt. \end{aligned}$$

There is a similar version for definite integrals.

Example 1.1 *Evaluate* $I = \int_0^1 \frac{dx}{\sqrt{4-2x-x^2}}$.

The integral is close to the integral $\int \frac{dx}{\sqrt{1-x^2}}$ which equals $\sin^{-1} x$ up to a constant, so we attempt to use this known integral. By completing square we may write

$$\begin{aligned} 4-2x-x^2 &= 5-(x+1)^2 \\ &= 5\left(1-\left(\frac{x+1}{\sqrt{5}}\right)^2\right), \end{aligned}$$

making substitution $t = \frac{x+1}{\sqrt{5}}$, $dx = \sqrt{5}dt$, where $t : \frac{1}{\sqrt{5}} \to \frac{2}{\sqrt{5}}$, we have

$$\begin{aligned} I &= \frac{1}{\sqrt{5}} \int_{\frac{1}{\sqrt{5}}}^{\frac{2}{\sqrt{5}}} \frac{\sqrt{5}dt}{\sqrt{1-t^2}} = \int_{\frac{1}{\sqrt{5}}}^{\frac{2}{\sqrt{5}}} \frac{dt}{\sqrt{1-t^2}} \\ &= \sin^{-1} \frac{2}{\sqrt{5}} - \sin^{-1} \frac{1}{\sqrt{5}}. \end{aligned}$$

Now let us recall the technique of integration by parts, which is in many aspects the soul of the analysis. Integration by parts is the integral form of the product rule for derivatives, since $(fg)' = f'g + fg'$ so that

$$f(x)g(x) = \int g(x)f'(x)dx + \int f(x)g'(x)dx$$

rearranging the terms to obtain

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx.$$

Similarly we have

$$\int_a^b f(x)g'(x)dx = f(x)g(x)|_a^b - \int_a^b g(x)f'(x)dx,$$

or in terms of differentials we can rewrite the preceding formula as

$$\int_a^b f(x)dg(x) = f(x)g(x)|_a^b - \int_a^b g(x)df(x).$$

However there is no general rule to tell us how to split an integrand into $g'(x)f(x)$.

Example 1.2 *Consider $I = \int xe^x dx$. Then*

$$\begin{aligned} I &= \int xde^x = xe^x - \int e^x dx \\ &= (x-1)e^x + C. \end{aligned}$$

Example 1.3 *Now let us consider $I_n = \int x^n e^x dx$ where $n = 1, 2, 3, \cdots$. Using integration by parts*

$$\begin{aligned} I_n &= \int x^n de^x = x^n e^x - \int e^x dx^n \\ &= x^n e^x - n \int x^{n-1} e^x dx \\ &= x^n e^x - nI_{n-1} \end{aligned}$$

which gives a reduction formula. By repeating the use of integration by parts, one can eventually work out the result. For example

$$I_2 = x^2e^x - 2I_1 = \left(x^2 - 2(x-1)\right)e^x + C$$

and

$$\begin{aligned} I_3 &= x^3e^x - 3I_2 \\ &= \left[x^3 - 3\left(x^2 - 2(x-1)\right)\right]e^x + C \end{aligned}$$

etc.

Example 1.4 *Consider $I_n = \int \cos^n x dx$ where n is a non-negative integer. Split the integrand $\cos^n x$ into $\cos^{n-1} x \cos x = \cos^{n-1} x(\sin' x)$, and perform integration by parts. Then*

$$\begin{aligned} I_n &= \int \cos^{n-1} x(\sin' x)dx = \int \cos^{n-1} x d\sin x \\ &= \cos^{n-1} x \sin x - \int \sin x d\cos^{n-1} x \\ &= \cos^{n-1} x \sin x - (n-1)\int \sin x \cos^{n-2} x(-\sin x)dx \\ &= \cos^{n-1} x \sin x + (n-1)\int \sin^2 x \cos^{n-2} x dx. \end{aligned}$$

Applying the identity $\sin^2 x = 1 - \cos^2 x$ in the last integral, we obtain

$$\begin{aligned} I_n &= \cos^{n-1} x \sin x + (n-1)\int (1-\cos^2 x)\cos^{n-2} x dx \\ &= \cos^{n-1} x \sin x + (n-1)I_{n-2} - (n-1)I_n. \end{aligned}$$

Collecting I_n together to obtain

$$nI_n = (n-1)I_{n-2} + \cos^{n-1} x \sin x$$

so that

$$I_n = \frac{n-1}{n}I_{n-2} + \frac{1}{n}\cos^{n-1} x \sin x$$

which reduces the calculation of I_n to I_0 or I_1, both are easy to evaluate. For example

$$\begin{aligned} \int_0^{\pi/2} \cos^n x dx &= \frac{n-1}{n}\int_0^{\pi/2} \cos^{n-2} x dx + \frac{1}{n}\cos^{n-1} x \sin x \Big|_0^{\pi/2} \\ &= \frac{n-1}{n}\int_0^{\pi/2} \cos^{n-2} x dx = \cdots \\ &= \begin{cases} \frac{n-1}{n}\frac{n-3}{n-2}\cdots\int_0^{\pi/2}\cos x dx & \textit{if } n = \textit{old}, \\ \frac{n-1}{n}\frac{n-3}{n-2}\cdots\int_0^{\pi/2} dx & \textit{if } n = \textit{even}. \end{cases} \end{aligned}$$

Example 1.5 *Consider* $I = \int e^x \sin x dx$. *We have*

$$\begin{aligned} I &= -\int e^x d\cos x = -e^x \cos x + \int e^x \cos x dx \\ &= -e^x \cos x + \int e^x d \sin x \\ &= -e^x \cos x + e^x \sin x - \int e^x \sin x dx \\ &= -e^x \cos x + e^x \sin x - I \end{aligned}$$

so that

$$2I = -e^x \cos x + e^x \sin x + C.$$

2 First order differential equations

A (ordinary) differential equation is an equation involving an independent variable x, a function $y(x)$ and its derivatives:

$$F(x, y, y', \cdots, y^{(n)}) = 0.$$

By solving the highest order derivative $y^{(n)}$ in terms of lower order derivatives $y^{(k)}$ for $k < n$ and x, the above equation may be written as

$$y^{(n)} = f(x, y, y', \cdots, y^{(n-1)}). \tag{2.1}$$

Such an equation is called an n-th order differential equation. If $n = 1$, then it is called a first order differential equation. Thus a first order differential equation has the general form $y' = f(x, y)$, or implicitly $F(x, y, y') = 0$.

A function $y = \varphi(x)$ defined on some interval J is called a solution of (2.1) if

$$\varphi^{(n)}(x) = f(x, \varphi(x), \varphi'(x), \cdots, \varphi^{(n-1)}(x)) \quad \forall x \in J.$$

A function $y = \varphi(x)$ which contains n independent arbitrary constants $C_1, \cdots, C_n$ is called the general solution of (2.1) if 1) it is a solution for any arbitrary choice of $C_1, \cdots, C_n$, 2) any solution of (2.1) has this form.

The concept of general solutions is not very useful. We are often interested in the so-called initial problems or boundary problems. Observe that in order to determine the constants $C_1, \cdots, C_n$ in general we need n conditions which appear as initial conditions. More precisely, an initial condition for n-th order differential equation (2.1) may be formulated as

$$y(x_0) = y_0, \cdots, y^{(n-1)}(x_0) = y_{n-1}$$

where $x_0 \in J$ and $y_0, \cdots, y_{n-1}$ are given data.

A differential equation is called an (inhomogeneous) linear differential equation, if it is linear in $y, y', \cdots, y^{(n)}$, so that a linear differential equation can be written as the following general form

$$a_n(x) y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_0(x) y = h(x)$$

where $a_n, \cdots, a_0$ and h are functions of x. If $h \equiv 0$, then the linear equation is homogenous.

A first order linear differential equation can be thus put in the following general form

$$y' + p(x) y = q(x).$$

2.1 Separable first order DE

Consider a first order differential equation $\frac{dy}{dx} = f(x, y)$. It is separable if $f(x, y) = a(x)b(y)$, so that $\frac{dy}{dx} = a(x)b(y)$. Dividing the equation by $b(y)$ and multiplying it by dx to separate the variables x and y and write the equation to be

$$\frac{dy}{b(y)} = a(x)dx.$$

Integrating both sides of the equation to obtain the solution given by

$$\int \frac{dy}{b(y)} = \int a(x)dx$$

which gives in general solutions of a separable equation implicitly. If y_0 is a root to $b(y) = 0$, then clearly the constant function $y = y_0$ is also a solution.

Example 2.1 *Find the general solution to*

$$x(y^2 - 1) + y(x^2 - 1)\frac{dy}{dx} = 0.$$

The equation is separable and can be rearranged as

$$\frac{xdx}{x^2 - 1} + \frac{ydy}{y^2 - 1} = 0.$$

After integration we obtain

$$\ln |x^2 - 1| + \ln |y^2 - 1| = C$$

(C is a constant), which can be put in the form

$$(x^2 - 1)(y^2 - 1) = C.$$

The constant functions $y = 1$ or $y = -1$ are solutions but are already included in the above general form with $C = 0$.

Example 2.2 *Find the solution to $(1 + e^x)yy' = e^x$ satisfying the initial condition that $y(0) = 1$. The equation is separable:*

$$ydy = \frac{e^x}{1 + e^x}dx.$$

After integration we obtain the general solution

$$\frac{1}{2}y^2 = \ln(1 + e^x) + C.$$

To match the initial condition, we set $x = 0$ and $y = 1$ in the general solution to determine the constant $C = \frac{1}{2} - \ln 2$, so that $\frac{1}{2}y^2 = \ln(1 + e^x) + \frac{1}{2} - \ln 2$. After simplification we have

$$y^2 = \ln\left[\frac{e}{4}(1 + e^x)^2\right].$$

Some differential equations of first order can be transformed by proper substitutions to separable equations.

Example 2.3 *Find the general solution to $y' = \sin(x + y + 1)$.*

Let $u(x) = x + y(x) + 1$ so that $u' = 1 + y'$. The original equation can be formulated a DE of u, namely $u' = 1 + \sin u$ which is separable. Dividing the equation by $1 + \sin u$ and write the equation as

$$\frac{du}{1 + \sin u} = dx.$$

Integrating the equation we obtain

$$\int \frac{du}{1 + \sin u} = \int dx.$$

Let us evaluate the integral on the left hand side.

$$\begin{aligned}
\int \frac{du}{1 + \sin u} &= \int \frac{(1 - \sin u)du}{(1 + \sin u)(1 - \sin u)} \\
&= \int \frac{(1 - \sin u)du}{1 - \sin^2 u} = \int \frac{(1 - \sin u)du}{\cos^2 u} \\
&= \int \frac{du}{\cos^2 u} - \int \frac{\sin u du}{\cos^2 u} \\
&= \int \frac{du}{\cos^2 u} + \int \frac{d \cos u}{\cos^2 u} = \tan u - \frac{1}{\cos u} + C.
\end{aligned}$$

Therefore

$$\tan u - \frac{1}{\cos u} = x + C$$

or in terms of y and x, the solution is given by

$$\tan(x + y + 1) - \frac{1}{\cos(x + y + 1)} = x + C$$

or

$$\sin(x + y + 1) - 1 = (x + C) \cos(x + y + 1).$$

We also have solutions $x + y + 1 = 2n\pi - \frac{\pi}{2}$ where n =integers.

2.2 Homogenous equations

Consider a first order differential equation $\frac{dy}{dx} = f(x, y)$. If the function $f(x, y)$ (of two variables) is homogenous, i.e. $f(x, y) = h(\frac{y}{x})$ where h is a function of one variable, then we can make a substitution $u(x) = \frac{y(x)}{x}$ so that $y = xu$. The product rule gives that $\frac{dy}{dx} = u + x\frac{du}{dx}$, and the equation may be written as

$$u + x\frac{du}{dx} = f(u)$$

which is separable.

Example 2.4 *Find general solutions to $xy' = \sqrt{x^2 - y^2} + y$. The equation, by dividing x both sides, is homogenous*

$$y' = \sqrt{1 - \left(\frac{y}{x}\right)^2} + \frac{y}{x}$$

so we make substitution $u = \frac{y}{x}$ and change equation to be

$$u + x\frac{du}{dx} = \sqrt{1-u^2} + u.$$

Rearrange the equation: $\frac{du}{\sqrt{1-u^2}} = \frac{dx}{x}$. *Integrating both sides to obtain*

$$\sin^{-1} u = \ln|x| + C$$

or in terms of y, *general solutions are given by* $\sin^{-1}(\frac{y}{x}) = \ln|x| + C$, *together with solutions* $\frac{y}{x} = 1$ *and* $\frac{y}{x} = -1$.

Some differential equations of first order can be transformed into homogenous ones by simple substitutions.

For example, consider the following type of first order differential equations

$$\frac{dy}{dx} = f\left(\frac{a_1x + b_1y + c_1}{a_2x + b_2y + c_2}\right).$$

If $c_1 = c_2 = 0$ then the equation is homogenous, so we consider the case that c_1 or c_2 does not vanish. If

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0$$

and $b_1 \neq 0$, then we make substitution $u(x) = a_1x + b_1y(x)$ to transform the equation to a separable one. For the case where

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0$$

we make translation $x = t + k$ and $y = z + l$ such that

$$\begin{cases} a_1k + b_1l + c_1 = 0, \\ a_2k + b_2l + c_2 = 0. \end{cases}$$

Consider t as a new independent variable, and z as a function of t, then

$$z(t) = y(x) - l = y(t + k) - l$$

therefore, by chain rule,

$$\frac{dz}{dt} = \frac{dy}{dx}.$$

The differential equation we are interested becomes

$$\frac{dz}{dt} = f\left(\frac{a_1t + b_1z}{a_2t + b_2z}\right)$$

which is homogenous.

Example 2.5 *Find the general solution to*

$$y' = 2\left(\frac{y+2}{x+y-1}\right)^2.$$

Solve the linear system

$$\begin{cases} l+2=0, \\ k+l-1=0 \end{cases}$$

to obtain $l=-2$ and $k=3$. Let $x=t+3$ and $y=z-2$. Then the differential equation can be written as

$$z' = 2\left(\frac{z}{t+z}\right)^2$$

which is homogenous. Now making standard substitution $u(t)=\frac{z(t)}{t}$, so that $z'=u+tu'$ and

$$u + t\frac{du}{dt} = \frac{2u^2}{(1+u)^2}$$

which is separable. Rearrange the equation

$$t\frac{du}{dt} = \frac{2u^2 - u(1+u)^2}{(1+u)^2} = -\frac{u(1+u^2)}{(1+u)^2}$$

and separate the variables to obtain

$$\frac{(1+u)^2}{u(1+u^2)}du = -\frac{dt}{t}. \tag{2.2}$$

Since

$$\begin{aligned} \int \frac{(1+u)^2}{u(1+u^2)}du &= \int \left(\frac{1}{u} + \frac{2}{1+u^2}\right) du \\ &= \ln|u| + 2\tan^{-1} u \end{aligned}$$

therefore, by integrating the equation (2.2) we obtain

$$\ln|u| + 2\tan^{-1} u = -\ln|t| + C.$$

In terms of x and y the general solution is given by

$$\ln|y+2| + 2\tan^{-1}\frac{y+2}{x-3} = C.$$

2.3 Linear differential equations of first order

Consider a linear differential equation of first order

$$\frac{dy}{dx} + p(x)y = q(x) \tag{2.3}$$

where p and q are two continuous functions. The corresponding homogenous equation $\frac{dz}{dx}+p(x)z=0$ is separable, and has the general solution

$$z(x) = Ce^{-\int p(x)dx}$$

where $\int p(x)dx$ is a primitive of $p(x)$, and C is an arbitrary constant. It follows that $z(x)e^{\int p(x)dx}$ is a constant, so that

$$\frac{d}{dx}\left(z(x)e^{\int p(x)dx}\right) = 0$$

which is in turn equivalent to the homogenous equation $z' + p(x)z = 0$.

Next we consider the inhomogeneous equation (2.3). The previous discussion suggests to consider the differential of $y(x)e^{\int p(x)dx}$, and by employing the product rule for derivatives, we obtain

$$\begin{aligned} \frac{d}{dx}\left(y(x)e^{\int p(x)dx}\right) &= e^{\int p(x)dx}\left(\frac{dy}{dx} + p(x)y\right) \\ &= q(x)e^{\int p(x)dx} \end{aligned} \tag{2.4}$$

so by integrating the equation both sides we obtain

$$ye^{\int p(x)dx} = \int q(x)e^{\int p(x)dx}dx + C$$

dividing by $e^{\int p(x)dx}$ the equality to obtain the general solution of (2.3)

$$y = e^{-\int p(x)dx}\left(\int q(x)e^{\int p(x)dx}dx + C\right). \tag{2.5}$$

The function $e^{\int p(x)dx}$ which is multiplied to y to form $ye^{\int p(x)dx}$ is called an *integrating factor* to the inhomogeneous equation (2.3).

We may describe the above procedure to obtain general solutions for first order linear differential equations as following, which includes an idea that can be applied to other different situations, thus are worthy of learning.

Observe that $z(x) = e^{-\int p(x)dx}$ is a non-trivial solution to the corresponding homogenous equation $z' + p(x)z = 0$, in order to obtain the general solution to the inhomogeneous one (2.3), we make use of the solution $z(x)$: making substitution

$$u(x) = \frac{y(x)}{z(x)} \tag{2.6}$$

(which is a *standard substitution* as long as z is a known function which has some thing to do with the differential equation we are interested. We will use this idea in several occasions later on), and turn (2.3) into a differential equation in u. Of course, according to the explicit form of $z(x)$ we have $u(x) = y(x)e^{\int p(x)dx}$ and (2.4) just says that

$$u' = q(x)e^{\int p(x)dx}$$

which can be integrated to obtain the solution u.

Example 2.6 *Solve differential equation $y' + 2xy = 2xe^{-x^2}$.*

First work out an integrating factor $r(x) = e^{\int 2xdx} = e^{x^2}$. Multiplying $r(x)$ both sides the equation we obtain

$$e^{x^2}y' + 2xe^{x^2}y = 2x$$

that is

$$\frac{d}{dx}e^{x^2}y = 2x.$$

After integration we have

$$e^{x^2}y = x^2 + C$$

so that $y = \left(x^2 + C\right)e^{-x^2}$ is the general solution.

Example 2.7 *Bernoulli's equation is a non-linear first order equation*

$$\frac{dy}{dx} + p(x)y = q(x)y^n$$

where $n \neq 0$ *or* 1 *(but not necessary an integer).*

Dividing by y^n, the equation becomes

$$\frac{1}{y^n}\frac{dy}{dx} + p(x)y^{1-n} = q(x)$$

By using transformation $z = y^{1-n}$ the equation is transformed to a linear equation

$$\frac{dz}{dx} + (1-n)p(x)z = (1-n)q(x)$$

so that

$$y^{1-n} = e^{-(1-n)\int p(x)dx}\left((1-n)\int q(x)e^{(1-n)\int p(x)dx}dx + C\right).$$

3 Linear differential equations

Differential equations of second order play a special role in science. Many physical equations are second order ordinary or partial differential equations, such as the dynamics described by Newton's law of gravity, fluid dynamics which are determined by the fluid equations: Navier-Stokes equations.

3.1 Structure of general solutions to linear differential equations

Let us first describe the structures of solutions to linear differential equations. Recall the general linear differential equation of order n is an equation that can be written

$$a_n(x)y^{(n)} + \cdots + a_1(x)y' + a_0(x)y = f(x) \tag{3.1}$$

where a_i are continuous functions (on some interval) and $a_n \neq 0$.

Suppose y_p is a *particular* solution of (3.1), then clearly, y is a solution to (3.1) if and only if $y - y_p$ is a solution to the *corresponding homogenous* linear DE of n-th order

$$a_n(x)y^{(n)} + \cdots + a_1(x)y' + a_0(x)y = 0. \tag{3.2}$$

If y_1 and y_2 are two solutions to (3.2), then so is $\lambda y_1 + \mu y_2$, and moreover, there are n *linearly independent* solutions $y_1, \cdots, y_n$ of (3.2) such that the general solution

$$y = C_1y_1 + \cdots + C_ny_n$$

where $C_1, \cdots, C_n$ are arbitrary constants. That is, the collection of all solutions to a homogenous linear equation of n-th order is a *vector space* of dimension n. It follows that the general solution to (3.1) is given by

$$y = C_1 y_1 + \cdots + C_n y_n + y_p$$

where y_p is a particular solution of (3.1), $C_1 y_1 + \cdots + C_n y_n$ is the general solution to the corresponding homogenous equation (3.2).

Let us investigate again the general observation we have used to solve general linear differential equation of first order. That is, if there is a non-trivial function $z(x)$ which has some connection to the differential equation we are interested (for example, for a linear equation, the function may be a solution to the corresponding homogenous equation), we can make use of the known function in a canonical way by making substitution that $u(x) = \frac{y(x)}{z(x)}$ and work with the differential equation that u must satisfy.

Obviously the constant zero function is a trivial solution to any homogenous linear equation which of course give us no additional information. Suppose however we know, say by inspection, a non-trivial solution $z(x)$ to the homogenous equation (3.2), then we may reduce the equation to a lower order differential equation. Let us demonstrate this idea for homogeneous second order differential equations, for simplicity.

Suppose $z(x) \neq 0$ is a non-trivial solution to a homogenous linear differential equation of second order

$$p(x)\frac{d^2y}{dx^2} + q(x)\frac{dy}{dx} + r(x)y = 0. \tag{3.3}$$

Making the standard substitution $u(x) = \frac{y(x)}{z(x)}$, so that $y = uz$. Then $y' = u'z + uz'$ and $y'' = u''z + 2u'z' + uz''$, substitute these equations to (3.3) we obtain

$$p(x)\left(u''z + 2u'z' + uz''\right) + q(x)\left(u'z + uz'\right) + r(x)uz = 0.$$

Rearrange the above equation and use the fact that z is a solution to (3.3)

$$p(x)z(x)\frac{d^2u}{dx^2} + \left(2p(x)z'(x) + q(x)z(x)\right)\frac{du}{dx} = 0 \tag{3.4}$$

which is a homogenous differential equation of first order for unknown function $\frac{du}{dx}$.

Example 3.1 *Verify that* $z(x) = \frac{1}{x}$ *is a solution to*

$$xy'' + 2(1-x)y' - 2y = 0$$

hence find its general solution.

Since $z' = -x^{-2}$ and $z'' = 2x^{-3}$ we can easily see that z is a solution. Making substitution $y(x) = \frac{1}{x}u(x)$ in the equation we obtain a differential equation for u:

$$x\frac{1}{x}\frac{d^2u}{dx^2} + \left(-2x^{-2}x + 2(1-x)\frac{1}{x}\right)\frac{du}{dx} = 0.$$

Let $w = \frac{du}{dx}$ and simplify the above equation:

$$\frac{dw}{dx} - 2w = 0$$

which is separable, and has the general solution $w(x) = C_1 e^{2x}$. Integrating w to obtain

$$u(x) = \int w(x)dx = C_1 e^{2x} + C_2$$

so that

$$y(x) = \frac{1}{x}\left(C_1 e^{2x} + C_2\right) \tag{3.5}$$

is the general solution, where C_1 and C_2 are arbitrary constants.

Example 3.2 *Find the general solution to the inhomogeneous linear equation*

$$xy'' + 2(1-x)y' - 2y = 12x.$$

We have found the general solution to the corresponding homogenous equation which is given by (3.5), thus, according to the structure of solutions to linear equations, we only need to find a particular solution. Since the coefficients of the equation are all polynomials in x so we may look for a solution with a form $y(x) = ax+b$ where a, b are constants. Plugging into the equation $y'' = 0$, $y' = a$ and $y = ax + b$ into the equation

$$2a(1-x) - 2(ax+b) = 12x$$

so we should have $2a - 2b = 0$ and $-2a - 2a = 12$ so that $a = -3$ and $b = -3$. Thus $y_0(x) = -3x - 3$ is a particular solution, and the general solution thus is given by

$$y(x) = \frac{1}{x}\left(C_1 e^{2x} + C_2\right) - 3x - 3.$$

3.2 Linear ODE with constant coefficients

For homogenous linear ODE with *constant coefficients*:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0 \tag{3.6}$$

where $a_{n-1}, \cdots, a_0$ are constants, we can construct its general solution if we can find the roots to the *auxiliary equation*

$$m^n + a_{n-1}m^{n-1} + \cdots + a_1 m + a_0 = 0. \tag{3.7}$$

The auxiliary equation comes from the following observation. Since the derivative of e^{mx} is me^{mx} it is thus reasonable to search for a solution $y = e^{mx}$. Substitute $y^{(k)} = m^k e^{mx}$ into (3.6) we have

$$\left(m^n + a_{n-1}m^{n-1} + \cdots + a_1 m + a_0\right)e^{mx} = 0$$

thus e^{mx} is a solution if and only if m is a root to (3.7) and as long as m is real. If $m = \alpha + \beta i$ is a complex root of the auxiliary equation, then since the coefficients $a_{n-1}, \cdots, a_0$ are real numbers, so that $\bar{m} = \alpha - \beta i$ is also a root. Now the complex functions e^{mx} and $e^{\bar{m}x}$ both satisfy the differential equation (3.6) so that the real part and imaginary parts of

$$e^{mx} = e^{\alpha x}\cos(\beta x) + ie^{\alpha x}\sin(\beta x)$$

(Euler's equation) are solutions of (3.6), i.e. if $m = \alpha + \beta i$ is a complex root of the auxiliary equation, then

$$y_1(x) = e^{\alpha x} \cos(\beta x)$$

and

$$y_2(x) = e^{\alpha x} \sin(\beta x)$$

are a pair of *linearly independent* solutions of (3.6).

If m is a repeated root to the auxiliary equation with multiplicity $k \geq 2$, then $e^{mx}, xe^{mx}, \cdots, x^{k-1}e^{mx}$ are solutions. The similar conclusion is valid for complex roots. We therefore are able to construct n linearly independent solutions to (3.6) via the roots to the auxiliary equation.

Example 3.3 *Consider the harmonic motion described by*

$$\frac{d^2y}{dx^2} + \omega^2 y = 0$$

where $\omega \neq 0$ is real. The auxiliary equation is $m^2 + \omega^2 = 0$ which has two complex roots $m = \omega i$ and $\bar{m}$. So we have two independent solutions $\cos \omega x$ and $\sin \omega x$ and the general solution

$$y(x) = A \cos \omega x + B \sin \omega x$$

where A, B are arbitrary constants.

Example 3.4 *Solve the equation*

$$\frac{d^3y}{dx^3} - 4\frac{d^2y}{dx^2} + \frac{dy}{dx} + 6y = 0.$$

The auxiliary equation

$$m^3 - 4m + m + 6 = 0$$

has roots $-1, 2, 3$ so the general solution

$$y(x) = C_1 e^{-x} + C_2 e^{2x} + C_3 e^{3x}.$$

The situation for second ED with constant coefficients is particularly simple. Consider the homogenous linear equation

$$\frac{d^2y}{dx^2} + a\frac{dy}{dx} + by = 0 \tag{3.8}$$

where a, b are two real numbers.

Theorem 3.5 *Suppose the auxiliary equation*

$$m^2 + am + b = 0$$

has two roots m_1 and m_2.

1) If $m_1 \neq m_2$ are real, then the general solution is given by

$$y(x) = C_1 e^{m_1 x} + C_2 e^{m_2 x} \ .$$

2) If $m = m_1 = m_2$ is a repeated real root, then the general solution

$$y(x) = (C_1 + C_2 x)\, e^{mx}.$$

3) If $m_1 = \alpha + i\beta$ is a complex root ($\beta \neq 0$) so that $m_2 = \alpha - i\beta$, then the general solution

$$y(x) = e^{\alpha x}\left(C_1 \cos \beta x + C_2 \sin \beta x\right).$$

Proof. Note that $a = -(m_1 + m_2)$ and $b = m_1 m_2$. We consider 1) and 2) first. In this case $e^{m_1 x}$ is a solution, so we make substitution $y(x) = u(x)e^{m_1 x}$ in the differential equation. Since

$$y' = \left(u' + m_1 u\right) e^{m_1 x}$$

and

$$y'' = \left(u'' + 2m_1 u' + m_1^2 u\right) e^{m_1 x}$$

we obtain

$$\left(u'' + 2m_1 u' + m_1^2 u\right) + a\left(\left(u' + m_1 u\right)\right) + bu = 0.$$

Using the fact that λ_1 is a root and that $a = -(m_1 + m_2)$, we have

$$u'' - (m_2 - m_1)\, u' = 0.$$

Thus, if $m_2 - m_1 \neq 0$,

$$u'(x) = C_1 e^{(m_2 - m_1)x}$$

and integrating the equation to obtain

$$u(x) = C_1 e^{(m_2 - m_1)x} + C_2$$

which proves 1). If $m_2 - m_1 = 0$ then $u'' = 0$ so by integrating twice to obtain

$$u(x) = C_1 + C_2 x$$

which shows 2). ■

Example 3.6 *Solve the differential equation*

$$\frac{d^2 y}{dx^2} - 2\frac{dy}{dx} + 5y = 0.$$

The auxiliary equation $m^2 - 2m + 5 = 0$ has complex root $m = 1 + 2i$ and $\bar{m}$, so the general solution

$$y(x) = C_1 e^x \cos 2x + C_2 e^x \sin 2x.$$

Next we give some examples for inhomogeneous linear equations.

Example 3.7 *Solve the equation*

$$\frac{d^2 y}{dx^2} + 4y = \sin 3x.$$

It is easy to find the general solution to the corresponding homogeneous equation

$$\frac{d^2y}{dx^2} + 4y = 0$$

whose auxiliary equation $m^2 + 4 = 0$ has two complex roots $\pm 2i$. Since $\sin 3x$ is the imaginary part of e^{3ix}, and $3i$ is not a root of the auxiliary equation. Thus we search for a particular solution $y_p(x) = A \sin 3x$. Plugging into the equation we find $A = -\frac{1}{5}$. Hence the general solution

$$y(x) = C_1 \cos 2x + C_2 \sin 2x - \frac{1}{5} \sin 3x.$$

Example 3.8 *Consider*

$$\frac{d^2y}{dx^2} + 4\frac{dy}{dx} + 4y = \sin 3x.$$

The auxiliary equation $m^2 + 4m + 4$ has a repeated root -2. There is no particular solution with a form $A \sin 3x$ by a simple inspection, instead we look for a particular solution

$$y_p(x) = A \cos 3x + B \sin 3x.$$

Then

$$-9A + 12B + 4A = 0$$

and

$$-9B - 12A + 4B = 1.$$

Thus

$$A = -\frac{12}{169},\ B = -\frac{5}{169}.$$

The general solution

$$y(x) = (C_1 x + C_2)\, e^{-2x} - \frac{12}{169} \cos 3x - \frac{5}{169} \sin 3x.$$

Example 3.9 *Let us now consider*

$$\frac{d^2y}{dx^2} + 4y = \sin 2x.$$

We have seen that $\sin 2x$ is a solution to the corresponding homogenous equation, so we look for a particular solution

$$y_p(x) = Ax \cos 2x + Bx \sin 2x.$$

Then $B = 0$ and $A = -\frac{1}{4}$, so the general solution

$$y(x) = C_1 \cos 2x + C_2 \sin 2x - \frac{1}{4} x \cos 2x.$$

Example 3.10 *Find a particular solution to*

$$\frac{d^2y}{dx^2} + 4y = \sin x + \sin 2x \ .$$

By a simple inspection, $y_1 = \frac{1}{3}\sin x$ *is a particular solution to*

$$\frac{d^2y}{dx^2} + 4y = \sin x$$

and we know from the previous example $y_2 = -\frac{1}{4}x\cos 2x$ *is a particular solution to*

$$\frac{d^2y}{dx^2} + 4y = \sin 2x.$$

Thus

$$y_p = \frac{1}{3}\sin x - \frac{1}{4}x\cos 2x$$

is a particular solution.

Example 3.11 *Let us consider inhomogeneous linear equation*

$$\frac{d^2y}{dx^2} - 3\frac{dy}{dx} + 2y = f(x)$$

where $f(x)$ *is a given function.*

The auxiliary equation $m^2 - 3m + 2 = 0$ has two real roots 1 and 2, so the general solution to the corresponding homogenous equation is $C_1e^x + C_2e^{2x}$.

1) Suppose $f(x) = \sin x$ which is the imaginary part of e^{ix}, since i is not a root of the auxiliary equation, so we may search for a particular solution $y_p = A\sin x + B\cos x$, but not just $A\sin x$ which is not good. Feeding y_p, $y_p' = A\cos x - B\sin x$ and $y_p'' = -y_p$ into the differential equation

$$-y_p - 3(A\cos x - B\sin x) + 2y_p = \sin x$$

and collecting the terms of $\sin x$ and $\cos x$ together to obtain

$$(2A + 3B - 1)\sin x + (2B - 3A)\cos x = 0.$$

Set $2A + 3B - 1 = 0$ and $2B - 3A = 0$, and solve the system to obtain $A = \frac{2}{13}$ and $B = \frac{3}{13}$. The general solution is given by

$$y = C_1e^x + C_2e^{2x} + \frac{2}{13}\sin x + \frac{3}{13}\cos x.$$

2) $f(x) = e^{3x}$. Since 3 is not a root of the auxiliary equation, so search for a particular solution $y_p = Ae^{3x}$. Feeding it into the differential equation:

$$(9A - 9A + 2A)e^{3x} = e^{3x}$$

to obtain a particular solution $y_p = \frac{1}{2}e^{3x}$.

If however $f(x) = e^{2x}$, then we may attempt a particular solution $y_p = Axe^{2x}$ as e^{2x} is a solution to the corresponding homogenous equation. Using the equations $y_p' = Ae^{2x} + 2y_p$ and

$$y_p'' = 2Ae^{2x} + 2y_p' = 4Ae^{2x} + 4y_p$$

feeding them into the differential equation

$$4Ae^{2x} + 4y_p - 3\left(Ae^{2x} + 2y_p\right) + 2y_p = e^{2x}$$

and collecting the terms e^{2x} and y_p together

$$(4A - 3A - 1)\, e^{2x} = 0$$

so that $A = 1$, i.e. $y_p = xe^{2x}$ is a solution, so that the general solution to

$$\frac{d^2y}{dx^2} - 3\frac{dy}{dx} + 2y = e^{2x}$$

is given by

$$y = C_1e^x + C_2e^{2x} + xe^{2x}$$

3) $f(x) = xe^{2x}$. Since 2 is a root of the auxiliary equation, so we may search for a particular solution in a form $y_p = (Ax^2 + Bx)e^{2x}$ (we have included Bxe^{2x} as well, since e^{2x} is a solution to the homogenous equation, but not xe^{2x}).

4) $f(x) = e^x \sin x$ which is the imaginary part of $e^{(1+i)x}$ and $1 + i$ is not a root of the auxiliary equation, so we may search for a particular solution $y_p = (A\cos x + B\sin x)e^x$.

5) $f(x) = \sin^2 x$. Since $\sin^2 x = \frac{1}{2} - \frac{1}{2}\cos 2x$, so we may attempt a particular solution with a form $y_p = A + B\cos 2x + C\sin 2x$.

4 Some facts about matrices

An $m \times n$ matrix A is an array of numbers arranged into m rows and n columns

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

and simply written as $A = (a_{ij})$, where a_{ij} is the entry in the ith row and jth column. If $m = n$, then A is called a square matrix.

Let us concentrate on 2×2 matrices. You will learn the general theory about matrices in linear algebra (topics in your paper Mathematics I).

First of all, we have elementary operations among 2×2 matrices: if

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},\ B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

then we can form a matrix $A+B$ by simply adding their corresponding entries

$$A+B=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix}.$$

If λ is a number we may form a matrix

$$\lambda A=\begin{pmatrix} \lambda a_{11} & \lambda a_{12} \\ \lambda a_{21} & \lambda a_{22} \end{pmatrix}.$$

That is $A \pm B = (a_{ij} \pm b_{ij})$ and $\lambda A = (\lambda a_{ij})$. The more interesting operation is the multiplication of two matrices, defined as the following

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} \end{pmatrix}. \end{aligned}$$

That is, if $AB=(c_{ij})$ then the entry

$$\begin{aligned} c_{ij} &= (a_{i1}, a_{i2})\begin{pmatrix} b_{1j} \\ b_{2j} \end{pmatrix} \\ &= \text{dot product of } (a_{i1}, a_{i2}) \text{ and } (b_{1j}, b_{2j}). \end{aligned}$$

Example 4.1 *Let*

$$A=\begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix} \text{ and } B=\begin{pmatrix} 0 & 3 \\ -2 & 5 \end{pmatrix}$$

Find $A+B, A-B$, $-A$, AB and BA.

In general we have $A+B=B+A$, $C(A+B)=AC+CB$, $(AB)C=A(BC)$, but the multiplication of matrices is in general not commutative.

The determinate of a 2×2 matrix A is denoted by $\det(A)$ or $|A|$ defined by

$$\det(A)=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}.$$

The mapping $A \to |A|$ is not additive, but it is multiplicative i.e.

$$\det(AB)=\det(A)\det(B)=\det(BA).$$

We will use I to denote the identity matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is trivial that $IA=AI$ for any 2×2 matrix. Clearly $|I|=1$.

Given a 2×2 matrix A, we say a 2×2 matrix B (if ever exists) is an inverse matrix of A if $AB=BA=I$. Since $\det(AB)=\det(A)\det(B)$, so that a necessary condition for the existence of an inverse matrix to A is that $\det(A)\neq 0$. It turns out this condition is also sufficient.

Theorem 4.2 *Let $A = (a_{ij})$ be a 2×2 matrix. Then A has an inverse matrix if and only if* $\det(A) \neq 0$. *In this case, the inverse matrix is unique and thus is denoted by A^{-1}, given by*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Proof. By a direct computation we can see that A^{-1} defined as above is an inverse matrix. If B is an inverse of A, then

$$B = B(AA^{-1}) = (BA)A^{-1} = IA^{-1} = A^{-1}$$

so the inverse matrix is unique. ■

We observe that $\det(A) = 0$, i.e. $a_{11}a_{22} = a_{21}a_{12}$ means two row vectors $\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$ and $\begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix}$ are proportional, that is, they are *linearly dependent.*

Let us consider $\mathbb{R}^2$ as the vector space of row vectors $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ (also consider as 2×1 matrix). Let $A = (a_{ij})$ be a 2×2 matrix. Then we associate A a linear mapping from $\mathbb{R}^2 \to \mathbb{R}^2$ denoted by A and defined by

$$A\mathbf{v} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 \\ a_{21}v_1 + a_{22}v_2 \end{pmatrix}.$$

Proposition 4.3 *Let $A = (a_{ij})$ be a 2×2 matrix.*

1) The linear system $A\mathbf{v} = 0$ has no zero solutions if and only if $\det(A) = 0$.

2) The linear system $A\mathbf{v} = \lambda\mathbf{v}$ has no trivial solution $\mathbf{v} \neq 0$ if and only if λ is an eigenvalue *of A, that is,* $\det(A - \lambda I) = 0$ *(which is called the characteristic equation of A). In this case, $\mathbf{v}$ is called an* eigenvector *(corresponding to the eigenvalue λ).*

A square matrix $A = (a_{ij})$ is diagonal if $a_{ij} = 0$ for any $i \neq j$.

Theorem 4.4 *Suppose a 2×2 matrix $A = (a_{ij})$ has distinct real eigenvalues λ_1 and λ_2, and let $\mathbf{v}_i = \begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix}$ be corresponding vectors with eigenvalues λ_i $(i = 1, 2)$. Let*

$$P = (\mathbf{v}_1, \mathbf{v}_2) = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}.$$

Then

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Proof. First show that P is invertible, which is equivalent to that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Suppose $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 = 0$, so that $\alpha A\mathbf{v}_1 + \beta A\mathbf{v}_2 = 0$. Hence $\alpha\lambda_1\mathbf{v}_1 + \beta\lambda_2\mathbf{v}_2 = 0$. It follows that $\beta(\lambda_2 - \lambda_1)\mathbf{v}_2 = 0$, so that $\beta = 0$ and similarly $\alpha = 0$. Therefore $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent, and P^{-1} exists.

By definition

$$\begin{aligned} AP &= A(\mathbf{v}_1, \mathbf{v}_2) = (A\mathbf{v}_1, A\mathbf{v}_2) \\ &= (\lambda_1\mathbf{v}_1, \lambda_2\mathbf{v}_2) = P \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \end{aligned}$$

Since P^{-1} exists so that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

■

Example 4.5 *Find all the eigenvalues and eigenvectors for the following matrices*

$$A = \begin{pmatrix} 2 & 1 \\ 6 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \text{ and } C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

5 Systems of linear differential equations

Consider a linear differential equation of order n:

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_0 y = f(t).$$

By introducing functions $y_k = y^{(k)}$ where $k = 0, \cdots, n-1$, the previous linear equation of order n is equivalent to the following system of linear equations of first order:

$$\begin{cases} y'_{n-1} &= -a_{n-1}y_{n-1} - \cdots - a_0 y_0 + f(t), \\ y'_{n-2} &= y_{n-1}, \\ \cdots & \\ \cdots & \\ y'_0 &= y_1. \end{cases}$$

For example, a second order linear differential equation

$$\frac{d^2y}{dt^2} + a\frac{dy}{dt} + by = f(t)$$

is equivalent to the system

$$\begin{cases} \frac{dx}{dt} &= -ax - by + f(t) \\ \frac{dy}{dt} &= x. \end{cases}$$

In terms of matrix notations, it can be written as

$$\left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} -a & -b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} f(t) \\ 0 \end{pmatrix}. \right.$$

Example 5.1 *Solve the following initial value problem*

$$\begin{cases} \frac{dx}{dt} &= 3x + y, \quad x(0) = 1; \\ \frac{dy}{dt} &= 6x + 4y, \quad y(0) = 1. \end{cases}$$

Method 1. From the first equation, substitute $y = \frac{dx}{dt} - 3x$ to the second equation, to obtain

$$\frac{d^2x}{dt^2} - 3\frac{dx}{dt} = 6x + 4\frac{dx}{dt} - 12x$$

so x solves the homogenous linear equation of second order

$$\frac{d^2x}{dt^2} - 7\frac{dx}{dt} + 6x = 0$$

whose auxiliary equation has roots 1 and 6, so $x(t) = C_1e^t + C_2e^{6t}$. Since $x(0) = 1$ and $x'(0) = y(0) + 3x(0) = 4$ we have

$$C_1 + C_2 = 1, \; C_1 + 6C_2 = 4.$$

Method 2. By chain rule (or the invariance of first order differentials) we have

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{6x+4y}{3x+y}$$

which is homogenous. By substitution $u = \frac{y}{x}$, then $y = xu$, $y' = u + xu'$, so that

$$u + x\frac{du}{dx} = \frac{6+4u}{3+u}$$

which is separable.

We next describe another method which is contained in the following

Theorem 5.2 *Consider the system of linear equations with constant coefficients*

$$\left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right. .$$

Suppose $A = (a_{ij})$ has two distinct eigenvalues λ_k with corresponding eigenvectors $\mathbf{v}_k = \begin{pmatrix} v_{1k} \\ v_{2k} \end{pmatrix}$ $(k = 1, 2)$. Then the general solution of the system is given by

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = C_1e^{\lambda_1 t}\mathbf{v}_1 + C_2e^{\lambda_2 t}\mathbf{v}_1$$

where C_1 and C_2 are arbitrary constants.

Proof. Let $P = (\mathbf{v}_1, \mathbf{v}_2)$. Then we have shown that

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Let

$$\mathbf{z}(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = P^{-1}\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}.$$

Then

$$\begin{aligned} \frac{d}{dt}\mathbf{z}(t) &= P^{-1}\begin{pmatrix} \frac{d}{dt}x(t) \\ \frac{d}{dt}y(t) \end{pmatrix} = P^{-1}A\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \\ &= P^{-1}AP\mathbf{z}(t) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix}. \end{aligned}$$

That is

$$z_1'(t) = \lambda_1 z_1(t) \quad \text{and } z_2'(t) = \lambda_2 z_2(t)$$

so that $z_k(t) = C_k e^{\lambda_k t}$ $(k = 1, 2)$. Hence

$$\begin{aligned} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} &= P \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = (\mathbf{v}_1, \mathbf{v}_2) \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} \\ &= C_1 e^{\lambda_1 t} \mathbf{v}_1 + C_2 e^{\lambda_2 t} \mathbf{v}_2. \end{aligned}$$

■

Remark 5.3 *If $\lambda_1 = \alpha + \beta i$ is complex (where $\beta \neq 0$) and $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 i$ is a (complex) eigenvector, then the general solution is given by*

$$\begin{aligned} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} &= e^{\alpha t} C_1 \left(\mathbf{v}_1 \cos \beta t - \mathbf{v}_2 \sin \beta t \right) \\ &\quad + e^{\alpha t} C_2 \left(\mathbf{v}_2 \cos \beta t + \mathbf{v}_1 \sin \beta t \right) \end{aligned}$$

where C_1, C_2 are arbitrary constants.

Example 5.4 *Solve the system of linear differential equations*

$$\left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right. .$$

Solve the characteristic equation $\det(A - \lambda I) = 0$, i.e. $\lambda^2 - 3\lambda + 2 = 0$ to obtain eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$. For $\lambda_1 = 1$, solve the linear system

$$\begin{pmatrix} 0-1 & 1 \\ -2 & 3-1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0$$

to obtain $c_1 = c_2$, so $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue 1. Similarly, solve the linear system

$$\begin{pmatrix} 0-2 & 1 \\ -2 & 3-2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0$$

to obtain an eigenvector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, so the general solution

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = C_1 e^t \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2 e^{2t} \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Example 5.5 *Solve the system*

$$\left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} 2 & -5 \\ 2 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right. .$$

The characteristic equation of the matrix in system

$$\begin{vmatrix} 2-\lambda & -5 \\ 2 & -4-\lambda \end{vmatrix} = \lambda^2 + 2\lambda + 2 = 0$$

has a pair conjugate complex roots $\lambda_1 = -1+i$ and $\lambda_2 = -1-i$. For $\lambda_1 = -1+i$ the linear system

$$\begin{pmatrix} 2-\lambda_1 & -5 \\ 2 & -4-\lambda_1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = 0$$

has a solution vector

$$\begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix} i$$

thus

$$\begin{aligned} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} &= e^{-t}C_1\left(\begin{pmatrix} 5 \\ 3 \end{pmatrix}\cos t - \begin{pmatrix} 0 \\ -1 \end{pmatrix}\sin t\right) \\ &\quad + C_2\left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}\cos t + \begin{pmatrix} 5 \\ 3 \end{pmatrix}\sin t\right). \end{aligned}$$

Example 5.6 *Solve the initial value problem to the linear system*

$$\left\{ \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -4 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \begin{matrix} x(0) = 1, \\ y(0) = 1. \end{matrix} \right.$$

The characteristic equation of the matrix in the system

$$\begin{vmatrix} 2-\lambda & 1 \\ -4 & 6-\lambda \end{vmatrix} = (\lambda-4)^2 = 0$$

has repeated root 4, so e^{4t} is a solution to the system. Taking into account the initial condition, we may set $x(t) = (At+1)e^{4t}$ and $y(t) = (Bt+1)e^{4t}$, and feed them into the system to obtain $A = -1$ and $B = -2$. Thus the solution to the initial problem is given by

$$\begin{cases} x(t) &= (1-t)e^{4t}, \\ y(t) &= (1-2t)e^{4t}. \end{cases}$$

6 Partial derivatives, chain rule

From this section, we study functions of several variables.

6.1 Computations of partial derivatives

Let us begin with a (real) function of two variables, $u = f(x, y)$ defined on an *open subset* such as an open disk, and begin with the partial derivatives of f. By saying a subset U of $\mathbb{R}^2$ (resp. $\mathbb{R}^n$) an open subset we mean that if any point $p \in U$ there is an open disk (resp. an open ball in $\mathbb{R}^n$) $B_p(r)$ centered at p with radius $r > 0$ such that $B_p(r) \subset U$.

Holding $y = y_0$ as constant, consider $f(x, y_0)$ as a function of x, if its derivative (in x) exists at x_0, i.e.

$$\lim_{x \to x_0} \frac{f(x, y_0) - f(x_0, y_0)}{x - x_0}$$

exists, then its limit is called the partial derivative of f in x at (x_0, y_0), denoted by one of the following notations

$$\frac{\partial u}{\partial x}, \ \frac{\partial f(x_0, y_0)}{\partial x}; \ u_x, \ f_x(x_0, y_0); \ D_x u, \ D_x f(x_0, y_0).$$

(It was C.G.J.Jacobi who first proposed to use symbol ∂ instead of d for partial derivatives). Similarly we may introduce partial derivative in y, denoted by $\frac{\partial u}{\partial y}$ etc. The definition of partial derivatives applies well to functions of three variables, and to functions of several variables.

Example 6.1 *Find partial derivatives for $u = y^x$. Holding y as constant then u is an exponential function and $\frac{\partial u}{\partial x} = y^x \ln y$, while if hold x as constant, it is a power function so that $\frac{\partial u}{\partial y} = xy^{x-1}$.*

Example 6.2 *Find partial derivatives for $u = \frac{x}{x^2+y^2+z^2}$. The results are*

$$\begin{aligned} \frac{\partial u}{\partial x} &= -\frac{2x^2}{(x^2 + y^2 + z^2)^2} + \frac{1}{x^2 + y^2 + z^2} \\ &= \frac{y^2 + z^2 - x^2}{(x^2 + y^2 + z^2)^2}, \end{aligned}$$

$$\frac{\partial u}{\partial y} = -\frac{2xy}{(x^2 + y^2 + z^2)^2}, \ \frac{\partial u}{\partial z} = -\frac{2xz}{(x^2 + y^2 + z^2)^2}.$$

Example 6.3 *Let $u = yf(x^2 - y^2)$ where $f(t)$ is a differentiable function whose derivative is denoted by $f'(t)$. Then*

$$\frac{1}{x}\frac{\partial u}{\partial x} + \frac{1}{y}\frac{\partial u}{\partial y} = \frac{u}{y^2}.$$

In fact, according to the chain rule we have

$$\frac{\partial u}{\partial x} = 2xyf'(x^2 - y^2), \ \frac{\partial u}{\partial y} = f(x^2 - y^2) - 2xyf'(x^2 - y^2)$$

so that

$$\frac{1}{x}\frac{\partial u}{\partial x} + \frac{1}{y}\frac{\partial u}{\partial y} = \frac{1}{y}f(x^2 - y^2) = \frac{u}{y^2}.$$

Suppose $u = f(x, y)$ whose partial derivatives $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ exist on an open subset, so that $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ are functions of variables x and y. Suppose that the partial derivative in x of $\frac{\partial u}{\partial x}$ exists, then $\frac{\partial \frac{\partial u}{\partial x}}{\partial x} = \frac{\partial}{\partial x}(\frac{\partial u}{\partial x})$ is called the second order partial derivative of u, denoted by $\frac{\partial^2 u}{\partial x^2}$ or by any of the following

$$\frac{\partial^2 f(x_0, y_0)}{\partial x^2}; u_{xx}, f_{xx}(x_0, y_0); D^2_{xx} u, D^2_{xx} f(x_0, y_0).$$

Similarly $\frac{\partial}{\partial y}(\frac{\partial u}{\partial x})$ is denoted by $\frac{\partial^2 u}{\partial y \partial x}$ etc. Higher order derivatives can be defined inductively.

Example 6.4 *Find partial derivatives of* $u = \tan^{-1}\frac{x}{y}$ *up to second order. In fact*

$$\frac{\partial u}{\partial x} = \frac{y}{x^2+y^2}, \quad \frac{\partial u}{\partial y} = -\frac{x}{x^2+y^2}$$

and

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{y}{x^2+y^2}\right) = -\frac{2xy}{(x^2+y^2)^2}, \\ \frac{\partial^2 u}{\partial x \partial y} &= \frac{\partial}{\partial y}\left(\frac{y}{x^2+y^2}\right) = \frac{x^2-y^2}{(x^2+y^2)^2} = \frac{\partial^2 u}{\partial y \partial x}, \\ \frac{\partial^2 u}{\partial y^2} &= \frac{\partial}{\partial y}\left(-\frac{x}{x^2+y^2}\right) = \frac{2xy}{(x^2+y^2)^2}. \end{aligned}$$

In particular, u solves the Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

We can carry on to find

$$\frac{\partial^2 u}{\partial x^2 \partial y} = \frac{\partial}{\partial y}\left(-\frac{2xy}{(x^2+y^2)^2}\right) = \frac{6xy^2-2x^3}{(x^2+y^2)^3} \text{ and etc.}$$

6.2 The chain rule

Let us concentrate on functions of two variables for simplicity, but what we are going to do can be generalized to functions of several variables with proper modifications.

Lemma 6.5 *Suppose that* $u = f(x, y)$ *defined on an open subset* U *has first order partial derivatives* $\frac{\partial z}{\partial x}$ *and* $\frac{\partial z}{\partial y}$ *which are continuous functions on* U*. Let* $(x_0, y_0) \in U$*,* $\Delta x = x - x_0$*,* $\Delta y = y - y_0$ *and* $\Delta u = f(x, y) - f(x_0, y_0)$*. Then*

$$\Delta u = \frac{\partial f(x_0, y_0)}{\partial x}\Delta x + \frac{\partial f(x_0, y_0)}{\partial y}\Delta y + \alpha \tag{6.1}$$

where the remainder α *depends on* (x_0, y_0) *and* (x, y) *and*

$$\lim_{(x,y)\to(x_0,y_0)} \frac{\alpha}{\sqrt{\Delta x^2 + \Delta y^2}} = 0 \ .$$

That is, α *is small in comparison with* Δx *and* Δy*, thus the main part of the increment* Δu *at* (x_0, y_0) *is*

$$\frac{\partial f(x_0, y_0)}{\partial x}\Delta x + \frac{\partial f(x_0, y_0)}{\partial y}\Delta y$$

which is linear in the increments $(\Delta x, \Delta y)$ *of independent variables, called the first order* differential *of* f *at* (x_0, y_0)*.*

Proof. By a simple inspection we can see easily that

$$\begin{aligned}\alpha &= \left(\frac{f(x,y)-f(x_0,y)}{\Delta x} - \frac{\partial f(x_0,y)}{\partial x}\right)\Delta x \\ &\quad + \left(\frac{f(x_0,y)-f(x_0,y_0)}{\Delta y} - \frac{\partial f(x_0,y_0)}{\partial y}\right)\Delta y \\ &\quad + \left(\frac{\partial f(x_0,y)}{\partial x} - \frac{\partial f(x_0,y_0)}{\partial x}\right)\Delta x\end{aligned}$$

where we have used the convention that, if $\Delta x = 0$ (resp. $\Delta y = 0$) then the term $f(x,y)-f(x_0,y) = 0$ (resp. $f(x_0,y)-f(x_0,y_0) = 0$) and the term $\frac{f(x,y)-f(x_0,y)}{\Delta x}$ (resp. $\frac{f(x_0,y)-f(x_0,y_0)}{\Delta y}$) is regarded as zero. Then, since

$$\frac{\Delta x}{\sqrt{\Delta x^2+\Delta y^2}}, \frac{\Delta y}{\sqrt{\Delta x^2+\Delta y^2}}$$

are bounded,

$$\begin{aligned}\frac{f(x,y)-f(x_0,y)}{\Delta x} - \frac{\partial f(x_0,y)}{\partial x} &\to 0, \\ \frac{f(x_0,y)-f(x_0,y_0)}{\Delta y} - \frac{\partial f(x_0,y_0)}{\partial y} &\to 0\end{aligned}$$

and

$$\frac{\partial f(x_0,y)}{\partial x} - \frac{\partial f(x_0,y_0)}{\partial x} \to 0$$

as $\sqrt{\Delta x^2+\Delta y^2} \to 0$, we obtain

$$\frac{\alpha}{\sqrt{\Delta x^2+\Delta y^2}} \to 0 \text{ as } \sqrt{\Delta x^2+\Delta y^2} \to 0.$$

■

The first order *differential* of $u = f(x,y)$, denoted by du or df, and therefore

$$df = \frac{\partial f(x,y)}{\partial x}dx + \frac{\partial f(x,y)}{\partial y}dy.$$

The function

$$z = f(x_0,y_0) + \frac{\partial f(x_0,y_0)}{\partial x}(x-x_0) + \frac{\partial f(x_0,y_0)}{\partial y}(y-y_0)$$

is the *linear approximation* of $z = f(x,y)$ near the point (x_0,y_0). The above linear equation in (x,y,z) represents the *tangent plane* to the surface graph of the function $z = f(x,y)$ at the point $(x_0,y_0,f(x_0,y_0))$. We will return to this topic in the following lectures.

Lemma 6.6 *(Chain rule for two variable functions) Suppose that $f(x,y)$ is function on an open subset $U \subset R^2$ with continuous partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, and suppose $x = \varphi(t)$ and $y = \psi(t)$ are two differentiable functions on an interval (a,b) such that $(\varphi(t),\psi(t)) \in U$ for every $t \in (a,b)$. Let $F(t) = f(x(t),y(t))$. Then F is differentiable on (a,b) and*

$$\frac{dF(t)}{dt} = \frac{\partial f}{\partial x}\frac{d\varphi(t)}{dt} + \frac{\partial f}{\partial y}\frac{d\psi(t)}{dt} \tag{6.2}$$

where the partials $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are evaluated at $x = \varphi(t)$ and $y = \psi(t)$.

Proof. For t_0 and t in (a,b) we have

$$F(t)-F(t_0)=\frac{\partial f}{\partial x}\Delta x+\frac{\partial f}{\partial y}\Delta y+\alpha$$

where

$$\Delta x=\varphi(t)-\varphi(t_0),\ \ \Delta y=\psi(t)-\psi(t_0).$$

Hence

$$\frac{F(t)-F(t_0)}{\Delta t}=\frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t}+\frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t}+\frac{\alpha}{\Delta t}.$$

Letting $\Delta t\to 0$, we obtain

$$\frac{\Delta x}{\Delta t}\to\varphi'(t_0),\ \ \frac{\Delta y}{\Delta t}\to\psi'(t_0),$$

$$\frac{\alpha}{\Delta t}=\frac{\alpha}{\sqrt{\Delta x^2+\Delta y^2}}\sqrt{\left(\frac{\Delta x}{\Delta t}\right)^2+\left(\frac{\Delta y}{\Delta t}\right)^2}\frac{|\Delta t|}{\Delta t}\to 0,$$

and therefore

$$\begin{aligned} F'(t_0) &= \lim_{t\to t_0}\frac{F(t)-F(t_0)}{\Delta t} \\ &= \frac{\partial f}{\partial x}\lim_{t\to t_0}\frac{\Delta x}{\Delta t}+\frac{\partial f}{\partial y}\lim_{t\to t_0}\frac{\Delta y}{\Delta t}+\lim_{t\to t_0}\frac{\alpha}{\Delta t} \\ &= \frac{\partial f}{\partial x}\frac{d\varphi(t_0)}{dt}+\frac{\partial f}{\partial y}\frac{d\psi(t_0)}{dt}. \end{aligned}$$

■

Suppose we make change of variables: $x=\varphi(s,t)$ and $y=\psi(s,t)$, assume that φ and ψ have continuous partial derivatives. Consider $F(s,t)=f(\varphi(s,t),\psi(s,t))$. Holding t as constant and applying the chain rule (6.2) to variable s we obtain

$$\frac{\partial F}{\partial s}=\frac{\partial f}{\partial x}\frac{\partial\varphi}{\partial s}+\frac{\partial f}{\partial y}\frac{\partial\psi}{\partial s}$$

and similarly

$$\frac{\partial F}{\partial t}=\frac{\partial f}{\partial x}\frac{\partial\varphi}{\partial t}+\frac{\partial f}{\partial y}\frac{\partial\psi}{\partial t}.$$

In terms of matrices, the chain rule may be put into a neat form

$$\left(\frac{\partial F}{\partial s},\frac{\partial F}{\partial t}\right)=\left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right)\begin{pmatrix}\frac{\partial\varphi}{\partial s} & \frac{\partial\varphi}{\partial t}\\ \frac{\partial\psi}{\partial s} & \frac{\partial\psi}{\partial t}\end{pmatrix},$$

the 2×2 matrix on the right hand side is called the *first order total derivative* (or the Jacobian matrix) of the transformation $x=\varphi(s,t)$ and $y=\psi(s,t)$, denoted by $D(\varphi,\psi)$.

If all involved functions have continuous higher order partial derivatives, then we may repeat the use of the chain rule.

Example 6.7 *Let $F(s,t) = f(\varphi(s,t), \psi(s,t))$. Evaluate $\frac{\partial^2 F}{\partial t \partial s}$.*

By definition $\frac{\partial^2 F}{\partial t \partial s} = \frac{\partial}{\partial s}\left(\frac{\partial F}{\partial t}\right)$, so that

$$\begin{aligned}
\frac{\partial^2 F}{\partial t \partial s} &= \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial x}\frac{\partial \varphi}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial \psi}{\partial t}\right) \\
&= \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial x}\frac{\partial \varphi}{\partial t}\right) + \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial y}\frac{\partial \psi}{\partial t}\right) \\
&= \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial x}\right)\frac{\partial \varphi}{\partial t} + \frac{\partial f}{\partial x}\frac{\partial^2 \varphi}{\partial t \partial s} \\
&\quad + \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial y}\right)\frac{\partial \varphi}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial^2 \psi}{\partial t \partial s} \quad \textit{[Product rule]} \\
&= \left(\frac{\partial^2 f}{\partial x^2}\frac{\partial \varphi}{\partial s} + \frac{\partial^2 f}{\partial x \partial y}\frac{\partial \psi}{\partial s}\right)\frac{\partial \varphi}{\partial t} + \frac{\partial f}{\partial x}\frac{\partial^2 \varphi}{\partial t \partial s} \; \textit{[chain rule to } \frac{\partial f}{\partial x}\textit{]} \\
&\quad + \left(\frac{\partial^2 f}{\partial y \partial x}\frac{\partial \varphi}{\partial s} + \frac{\partial^2 f}{\partial y^2}\frac{\partial \psi}{\partial s}\right)\frac{\partial \psi}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial^2 \psi}{\partial t \partial s} \quad \textit{[chain rule to } \frac{\partial f}{\partial y}\textit{]}.
\end{aligned}$$

Here, the important thing we should keep in mind when working with higher partial derivatives is that the symbols $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are again functions of x and y, hence of s and t, so we have to apply the chain rule to these functions again.

As a direct consequence of the chain rule, we can show that the first order differentials are invariant under substitutions. To be more precise, if $F(s,t) = f(\varphi(s,t), \psi(s,t))$, then

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = \frac{\partial F}{\partial s}ds + \frac{\partial F}{\partial t}dt$$

where

$$dx = \frac{\partial \varphi}{\partial s}ds + \frac{\partial \varphi}{\partial t}dt, \quad dy = \frac{\partial \psi}{\partial s}ds + \frac{\partial \psi}{\partial t}dt.$$

The invariance of the first differentials under change of variables is useful in evaluating partial derivatives, but more importantly, it implies that differentials of functions are globally defined objects which do not depend on the coordinates we use to evaluate them.

Let us write down the chain rule for several variable functions.

Suppose that $f(x_1, \cdots, x_m)$ is a function of m variables which has continuous partial derivatives. Consider change of variables given by

$$\left\{\begin{array}{ccc} x_1 & = & \varphi_1(t_1, \cdots, t_n), \\ \vdots & \cdots & \vdots \\ x_m & = & \varphi_m(t_1, \cdots, t_n) \end{array}\right. \tag{6.3}$$

where $n \in \mathbb{N}$ and $\varphi_1, \cdots, \varphi_m$ are functions of $(t_1, \cdots, t_n)$ which have continuous derivatives $\frac{\partial \varphi_i}{\partial t_j}$. Let

$$F(t_1, \cdots, t_n) = f(\varphi_1(t_1, \cdots, t_n), \cdots, \varphi_m(t_1, \cdots, t_n)).$$

Then

$$\frac{\partial F}{\partial t_j} = \frac{\partial f}{\partial x_1}\frac{\partial \varphi_1}{\partial t_j} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial \varphi_m}{\partial t_j} \tag{6.4}$$

where $j = 1, \cdots, n$. In terms of matrix notations, the chain rule may be put in the following form

$$\left(\frac{\partial F}{\partial t_1}, \cdots, \frac{\partial F}{\partial t_n}\right) = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_m}\right) \begin{pmatrix} \frac{\partial \varphi_1}{\partial t_1} & \cdots & \frac{\partial \varphi_1}{\partial t_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_m}{\partial t_1} & \cdots & \frac{\partial \varphi_m}{\partial t_n} \end{pmatrix} \tag{6.5}$$

where the $m \times n$ matrix on the right-hand side of (6.5) is called the first order total derivative associated with the transformation (6.3), denoted by $D(\varphi_1, \cdots, \varphi_m)$. A careful study about the total derivatives for vector valued functions such as (6.3) will be the topic of Part A Option Multi-Variable Calculus.

Example 6.8 *Consider $u = x^y$ where $x = \varphi(t)$ and $y = \psi(t)$ so that $u(t) = \varphi(t)^{\psi(t)}$. According to the chain rule*

$$\begin{aligned} u'(t) &= \frac{\partial z}{\partial x}x'(t) + \frac{\partial z}{\partial y}y'(t) \\ &= \varphi'(t)yx^{y-1} + \psi'(t)x^y \ln x \\ &= \varphi'\psi\varphi^{\psi-1} + \psi'\varphi^{\psi} \ln \varphi. \end{aligned}$$

Example 6.9 *Let $u = f(x, y, z)$ have continuous partial derivatives. Let $x = \eta - \zeta$, $y = \zeta - \xi$ and $z = \xi - \eta$. Work out the matrix of the first order total derivative for the transformation*

$$\begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} & \frac{\partial x}{\partial \zeta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} & \frac{\partial y}{\partial \zeta} \\ \frac{\partial z}{\partial \xi} & \frac{\partial z}{\partial \eta} & \frac{\partial z}{\partial \zeta} \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

so, by the chain rule,

$$\begin{aligned} \left(\frac{\partial f}{\partial \xi}, \frac{\partial f}{\partial \eta}, \frac{\partial f}{\partial \zeta}\right) &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \\ &= \left(-\frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}, \frac{\partial f}{\partial x} - \frac{\partial f}{\partial z}, -\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\right), \end{aligned}$$

that is,

$$\frac{\partial u}{\partial \xi} = -\frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}, \quad \frac{\partial u}{\partial \eta} = \frac{\partial f}{\partial x} - \frac{\partial f}{\partial z}, \quad \frac{\partial u}{\partial \zeta} = -\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}.$$

Using chain rule again we have

$$\begin{aligned} \frac{\partial^2 u}{\partial \xi \partial \eta} &= \frac{\partial}{\partial \eta}\left(-\frac{\partial f}{\partial y} + \frac{\partial f}{\partial z}\right) = -\frac{\partial}{\partial \eta}\frac{\partial f}{\partial y} + \frac{\partial}{\partial \eta}\frac{\partial f}{\partial z} \\ &= -\left(\frac{\partial^2 f}{\partial y \partial x} - \frac{\partial^2 f}{\partial y \partial z}\right) + \frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial z \partial z}. \end{aligned}$$

6.3 Partial derivatives for implicit functions

The chain rule allows us to evaluate partial derivatives for implicit functions. Let us look at an example first.

Example 6.10 *Let $y = y(x)$ be the function implicitly given by the equation $x^2 + y^2 = 1$ where $x > 0$ and $y > 0$. Of course by solving y to obtain $y = \sqrt{1-x^2}$, thus*

$$\frac{dy}{dx} = -\frac{1}{2}\frac{2x}{\sqrt{1-x^2}} = -\frac{x}{\sqrt{1-x^2}}.$$

We can work out the derivative $\frac{dy}{dx}$ by just use the equation $x^2 + y^2 = 1$. Taking derivative both sides of the equation in x, keeping in mind y is a function of x, we obtain

$$\frac{d}{dx}\left(x^2 + y^2\right) = \frac{d}{dx}1 = 0$$

so that

$$2x + 2y\frac{dy}{dx} = 0$$

and solve $\frac{dy}{dx}$ to obtain $\frac{dy}{dx} = -\frac{x}{y}$ which gives just the same answer.

The idea used in the previous example can be applied to evaluating partial derivatives for implicit functions. Suppose that $y = y(x)$ is a function of x implicitly given by an equation

$$F(x, y) = 0 \ .$$

In order to solve y from the equation to determine the function $y = y(x)$ at least locally, we need to impose some conditions. Let us assume the partial derivatives of F (by considering x, y as independent variables) exist and are continuous, and assume that $\frac{\partial F}{\partial y} \neq 0$. To find out the derivative $\frac{dy}{dx}$, we take derivative both sides of the equation $F(x, y) = 0$, in x, keep in mind that $y = y(x)$ is a function of x. Then

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0.$$

The left hand side is the result from applying the chain rule to F with $x = x$ and $y = y(x)$. Since$F_y \neq 0$, by solving $\frac{dy}{dx}$ we obtain

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$

This idea applies to several variable implicit functions. For exam ple, if $z = z(x, y)$ is function implicitly given by the following equation

$$F(x, y, z) = 0,$$

and if, the partial derivatives of F (considering x, y, z as independent variables) are continuous, and $F_z \neq 0$, then, by taking derivative in x holding y as constant to obtain

$$F_x + F_z\frac{\partial z}{\partial x} = 0 \tag{6.6}$$

so that

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}.$$

Similarly we have

$$\frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

To compute the second partial derivatives, we continue the same procedure. Taking derivative both side of (6.6) in y to obtain

$$F_{xy} + F_{xz}\frac{\partial z}{\partial y} + \left(F_{zy} + F_{zz}\frac{\partial z}{\partial y}\right)\frac{\partial z}{\partial x} + F_z\frac{\partial^2 z}{\partial x \partial y} = 0$$

and solving $\frac{\partial^2 z}{\partial x \partial y}$ we obtain

$$\frac{\partial^2 z}{\partial x \partial y} = -\frac{F_{xy} + F_{xz}\frac{\partial z}{\partial y} + \left(F_{zy} + F_{zz}\frac{\partial z}{\partial y}\right)\frac{\partial z}{\partial x}}{F_z}$$

etc., though the formulate become increasingly complicated.

Finally we mention that the same idea applies to several functions with several variables. For example, from the following system

$$\begin{cases} F(x,y,z) = 0, \\ G(x,y,z) = 0, \end{cases} \tag{6.7}$$

we hope to solve y and z in terms of variable x, thus $y = y(x)$ and $z = z(x)$. By saying that $y(x)$ and $z(x)$ are solutions means that if we substitute (y, z) in the system (6.7) by $(y(x), z(x))$, then

$$F(x, y(x), z(x)) = 0, \ \ G(x, y(x), z(x)) = 0 \tag{6.8}$$

hold identically over the range of x. Therefore, by taking derivative on both sides of the equations in x, and employing the chain rule, we have

$$F_x + F_y\frac{dy}{dx} + F_z\frac{dz}{dx} = 0, \ G_x + G_y\frac{dy}{dx} + G_z\frac{dz}{dx} = 0, \tag{6.9}$$

which is a linear system in $(\frac{dy}{dx}\frac{dz}{dx})$, and can be put in a matrix form, namely

$$\begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix}\begin{pmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{pmatrix} = -\begin{pmatrix} F_x \\ G_x \end{pmatrix}. \tag{6.10}$$

We may solve $\frac{dy}{dx}$ and $\frac{dz}{dx}$ as long as

$$\det\begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix} = F_yG_z - F_zG_y \neq 0. \tag{6.11}$$

[In fact, near the point (x, y, z) where $F_yG_z - F_zG_y \neq 0$, we can show that (y, z) can be solved from the system (6.7) at least locally, which is a part of the conclusion of the so-called Inverse Function Theorem. The proper formulation and its proof of the inverse function theorem will be the topics

for the Part A option Multi-Variable Calculus in Hilary term]. Indeed, since under the condition (6.11) the matrix

$$\begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix}$$

is invertible, so that

$$\begin{aligned} \begin{pmatrix} \frac{dy}{dx} \\ \frac{dz}{dx} \end{pmatrix} &= -\begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix}^{-1} \begin{pmatrix} F_x \\ G_x \end{pmatrix} \\ &= -\frac{1}{F_yG_z - F_zG_y} \begin{pmatrix} G_z & -F_z \\ -G_y & F_y \end{pmatrix} \begin{pmatrix} F_x \\ G_x \end{pmatrix} \\ &= -\frac{1}{F_yG_z - F_zG_y} \begin{pmatrix} G_zF_x - G_xF_z \\ F_yG_x - F_xG_y \end{pmatrix}. \end{aligned}$$

Hence

$$\frac{dy}{dx} = -\frac{G_zF_x - G_xF_z}{F_yG_z - F_zG_y} = -\begin{vmatrix} F_z & F_x \\ G_z & G_x \end{vmatrix} \Big/ \begin{vmatrix} F_z & F_y \\ G_z & G_y \end{vmatrix}$$

and

$$\frac{dz}{dx} = -\frac{F_yG_x - F_xG_y}{F_yG_z - F_zG_y} = -\begin{vmatrix} F_y & F_x \\ G_y & G_x \end{vmatrix} \Big/ \begin{vmatrix} F_z & F_y \\ G_z & G_y \end{vmatrix}.$$

Example 6.11 *Let $y = y(x)$ and $z = z(x)$ be the functions satisfying the following equations*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad \text{and} \quad x + y + z = 0$$

where $x > 0$, $y > 0$ and $z > 0$. Find the derivatives $\frac{dy}{dx}$ and $\frac{dz}{dx}$.

We may differentiate the equations in x, while keep in mind y and z are functions of x, so according to chain rule

$$\frac{2x}{a^2} + \frac{2y}{b^2}\frac{dy}{dx} + \frac{2z}{c^2}\frac{dz}{dx} = \frac{d}{dx}1 = 0, \tag{6.12}$$

and

$$1 + \frac{dy}{dx} + \frac{dz}{dx} = \frac{d}{dx}0 = 0. \tag{6.13}$$

From (6.13) we obtain

$$\frac{dz}{dx} = -\frac{dy}{dx} - 1$$

and substitute it into (6.12) to get

$$\frac{x}{a^2} + \frac{y}{b^2}\frac{dy}{dx} + \frac{z}{c^2}\left(-\frac{dy}{dx} - 1\right) = 0$$

from which we may solve $\frac{dy}{dx}$, hence

$$\frac{dy}{dx} = \frac{\frac{z}{c^2} - \frac{x}{a^2}}{\frac{y}{b^2} - \frac{z}{c^2}}, \ \frac{dz}{dx} = \frac{\frac{y}{b^2} - \frac{x}{a^2}}{\frac{z}{c^2} - \frac{y}{b^2}}$$

at those points where $\frac{y}{b^2} - \frac{z}{c^2} \neq 0$.

6.4 Some differential operators

The symbols for partial differentiation such as $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ etc. may be considered as operations acting on functions (which have continuous partial derivatives), sending f to its partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ etc. It is useful to be familiar with some *differential operators* which are used extensively in science.

The symbol ∇ In the n-dimensional Euclidean space $\mathbb{R}^n$, the symbol ∇ means the total differentiation. Under the Cartesian coordinate system $(x_1, \cdots, x_n)$, ∇ denotes the total derivative

$$\nabla = \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right).$$

When ∇ applies to a function $f(x_1, \cdots, x_n)$ with continuous partial derivatives, ∇f means the total derivative

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)$$

called the *gradient (vector field)* of f. ∇f may be considered a function taking values in $\mathbb{R}^n$ (such a function is called a vector-valued function, also called a *vector field* in $\mathbb{R}^n$ in this special case that the number of functions in ∇f is exactly the dimension n of $\mathbb{R}^n$).

On the other hand, if

$$\mathbf{u}(x_1, \cdots, x_n) = (u^1(x_1, \cdots, x_n), \cdots, u^n(x_1, \cdots, x_n))$$

is a function of n variables defined on $U \subset \mathbb{R}^n$, taking values in $\mathbb{R}^n$ (a *vector field* on U) then we may make *dot product* between ∇ and $\mathbf{u}$ to obtain a real valued function by

$$\begin{aligned} \nabla \cdot \mathbf{u} &= \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right) \cdot \left(u^1, \cdots, u^n \right) \\ &= \frac{\partial u^1}{\partial x_1} + \cdots + \frac{\partial u^n}{\partial x_n} \end{aligned}$$

which is called the *divergence* of the vector field $\mathbf{u}$.

We have seen that, if $f(x_1, \cdots, x_n)$ is a scalar function on $U \subset \mathbb{R}^n$, then its gradient ∇f is a vector field, so we may apply dot product between ∇ and ∇f, to obtain

$$\begin{aligned} \nabla \cdot \nabla f &= \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right) \cdot \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right) \\ &= \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} \end{aligned}$$

which is called the *Laplacian* of f, denoted by Δf. Thus, we introduce the *differential operator* of second order

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

called the *Laplace operator* in $\mathbb{R}^n$. We extend the operation of Δ to vector valued functions as the following. Suppose

$$\mathbf{f}(x_1, \cdots, x_n) = \left(f^1(x_1, \cdots, x_n), \cdots, f^m(x_1, \cdots, x_n) \right)$$

is a vector valued function (where $m \in \mathbb{N}$) defined on $U \subset \mathbb{R}^n$, then we define

$$\Delta \mathbf{f} = \left(\Delta f^1, \cdots, \Delta f^m\right).$$

Curl operator $\nabla \times$ In 3-dimensional Euclidean space $\mathbb{R}^3$, besides the dot product, there is another multiplication between vectors called *cross products.* Recall that, under the Cartesian coordinate system (x, y, z), if $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ then the cross product $\mathbf{a} \times \mathbf{b}$ is defined by

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{k} \\ &= \left(a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1\right), \end{aligned}$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the standard basis in $\mathbb{R}^3$: $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$ and $\mathbf{k} = (0, 0, 1)$. $\mathbf{a} \times \mathbf{b}$ is the unique vector which is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ obeying the right hand rule, with magnitude $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \angle(\mathbf{a}, \mathbf{b})$, where $0 \leq \angle(\mathbf{a}, \mathbf{b}) \leq \pi$ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

We apply this definition by replacing $\mathbf{a}$ with $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ and a vector field $\mathbf{u} = (u^1, u^2, u^3)$ where u^1, u^2, u^3 are functions on $U \subset \mathbb{R}^3$ with continuous partial derivatives, and define the *curl* of the vector field $\mathbf{u}$ by

$$\begin{aligned} \nabla \times \mathbf{u} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u^1 & u^2 & u^3 \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u^2 & u^3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial z} \\ u^1 & u^3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \\ u^1 & u^2 \end{vmatrix} \mathbf{k} \\ &= \left(\frac{\partial u^3}{\partial y} - \frac{\partial u^2}{\partial z}, \frac{\partial u^1}{\partial z} - \frac{\partial u^3}{\partial x}, \frac{\partial u^2}{\partial x} - \frac{\partial u^1}{\partial y}\right). \end{aligned}$$

$\nabla \times \mathbf{u}$ is again a vector field on $U \subset \mathbb{R}^3$, also called the *vorticity* of $\mathbf{u}$.

Example 6.12 *Let $f(x, y, z) = x^2 + y^2 + z^2$. Compute*

$$\nabla f = 2(x, y, z)$$

and

$$\Delta f = 2(1 + 1 + 1) = 6$$

a constant function.

6.5 Change of coordinates and Jacobians

Sometimes it is useful to choose a special coordinate system which suites better to a specific problem. Suppose (x, y) (respectively (x, y, z) in $\mathbb{R}^3$) the Cartesian coordinate system in $\mathbb{R}^2$ (resp. in $\mathbb{R}^3$). Consider another coordinate system (u, v) which are given by equations $u = u(x, y)$ and $v = v(x, y)$, and equivalently $x = x(u, v)$ and $y = y(u, v)$. We call the mapping $(x, y) \to (u, v)$ a transformation of coordinates, or change of variables. We only consider those transformations

which have continuous partial derivatives. According to chain rule, if $f(x, y)$ is a function with continuous partial derivatives, then

$$\left(\frac{\partial f}{\partial u}, \frac{\partial f}{\partial v}\right) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

The determinate of the first order total derivative (the Jacobian matrix)

$$\det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

is called the Jacobian of the transformation, which will be denoted by $\frac{\partial(x,y)}{\partial(u,v)}$, i.e.

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

which is the *density of area elements* in a new coordinate system (u, v) in the following sense. Suppose the transformation $u = u(x, y)$ and $v = v(x, y)$ which send a domain U in xy-plane one to one and onto a domain D in uv-plane, then

$$\int_U f(x, y)dxdy = \int_D f(x(u, v), y(u, v)) \left|\frac{\partial(x, y)}{\partial(u, v)}\right| dudv.$$

That is to say, under the transformation $(x, y) \to (u, v)$, the area element $dxdy$ in the xy-plane is equivalent to $\left|\frac{\partial(x,y)}{\partial(u,v)}\right| dudv$, where $dudv$ is the area element in uv-plane.

Example 6.13 *(Parabolic coordinate system) The coordinates (u, v) given by the following relations*

$$x = \frac{1}{2}(u^2 - v^2), \quad y = uv$$

are called the parabolic coordinates in the planer. The Jacobian matrix and Jacobian are given by

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} u & -v \\ v & u \end{pmatrix}, \quad \frac{\partial(x, y)}{\partial(u, v)} = u^2 + v^2.$$

The transformation $(x, y) \to (u, v)$ is conformal *in the sense that*

$$\begin{aligned} (dx)^2 + (dy)^2 &= (udu - vdv)^2 + (vdu + udv)^2 \\ &= (u^2 + v^2)\left((du)^2 + (dv)^2\right). \end{aligned}$$

and

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = (u^2 + v^2)\left(\frac{\partial^2}{\partial u^2} + \frac{\partial^2}{\partial v^2}\right).$$

6.5.1 Polar coordinate system

If $P=(x,y)\in\mathbb{R}^2$ and $(x,y)\neq(0,0)$, then we can determinate the position of (x,y) by its distance r to $(0,0)$ and the angle θ from x-axis to $\overrightarrow{OP}$, so that $x=r\cos\theta$ and $y=r\sin\theta$. Then the Jacobian matrix

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

so that its Jacobian $\frac{\partial(x,y)}{\partial(r,\theta)}=r$.

On the other hand, $r=\sqrt{x^2+y^2}$ and $\tan\theta=\frac{y}{x}$. The Jacobian matrix of this transformation is given by

$$\begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}$$

so that its Jacobian $\frac{\partial(r,\theta)}{\partial(x,y)}=\frac{1}{\sqrt{x^2+y^2}}=\frac{1}{r}$. Hence

$$\frac{\partial(x,y)}{\partial(r,\theta)}\frac{\partial(r,\theta)}{\partial(x,y)}=1.$$

If $f(x,y)$ is a function with continuous partial derivatives, and $F(r,\theta)=f(r\cos\theta,r\sin\theta)$, then

$$\left\{\begin{array}{rcl} \frac{\partial F}{\partial r} & = & \cos\theta\frac{\partial f}{\partial x}+\sin\theta\frac{\partial f}{\partial y}, \\ \frac{\partial F}{\partial \theta} & = & -r\sin\theta\frac{\partial f}{\partial x}+r\cos\theta\frac{\partial f}{\partial y}. \end{array}\right. \tag{6.14}$$

It is also useful to express $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ in terms of $\frac{\partial f}{\partial r}$ and $\frac{\partial f}{\partial \theta}$, which can be achieved by solving $\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}$ from the above linear system:

$$\begin{aligned} \left(\frac{\partial f}{\partial x},\frac{\partial f}{\partial y}\right) &= \left(\frac{\partial f}{\partial r},\frac{\partial f}{\partial \theta}\right)\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}^{-1} \\ &= \frac{1}{r}\left(\frac{\partial f}{\partial r},\frac{\partial f}{\partial \theta}\right)\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \left(\frac{1}{r}\frac{\partial f}{\partial r},\frac{1}{r}\frac{\partial f}{\partial \theta}\right)\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \left(\cos\theta\frac{\partial f}{\partial r}-\frac{1}{r}\sin\theta\frac{\partial f}{\partial \theta},\sin\theta\frac{\partial f}{\partial r}+\frac{1}{r}\cos\theta\frac{\partial f}{\partial \theta}\right) \end{aligned}$$

that is

$$\left\{\begin{array}{rcl} \frac{\partial f}{\partial x} & = & \cos\theta\frac{\partial f}{\partial r}-\frac{\sin\theta}{r}\frac{\partial f}{\partial \theta}, \\ \frac{\partial f}{\partial y} & = & \sin\theta\frac{\partial f}{\partial r}+\frac{\cos\theta}{r}\frac{\partial f}{\partial \theta}. \end{array}\right. \tag{6.15}$$

It follows directly from (6.15) that

$$\left(\frac{\partial f}{\partial x}\right)^2+\left(\frac{\partial f}{\partial y}\right)^2=\left(\frac{\partial f}{\partial r}\right)^2+\frac{1}{r^2}\left(\frac{\partial f}{\partial \theta}\right)^2. \tag{6.16}$$

If $f(x, y)$ is a function with continuous partial derivatives up to order 2, then

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

We wish to work out the Laplace operator in the polar coordinate system. To this end, we continue to compute the second order partial derivatives. In fact,

$$\begin{aligned}
\frac{\partial^2 f}{\partial r^2} &= \frac{\partial}{\partial r}\left(\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}\right) \\
&= \cos\theta\frac{\partial}{\partial r}\frac{\partial f}{\partial x} + \sin\theta\frac{\partial}{\partial r}\frac{\partial f}{\partial y} \\
&= \cos\theta\left[\cos\theta\frac{\partial^2 f}{\partial x^2} + \sin\theta\frac{\partial^2 f}{\partial y\partial x}\right] \\
&\quad + \sin\theta\left[\cos\theta\frac{\partial^2 f}{\partial x\partial y} + \sin\theta\frac{\partial^2 f}{\partial y^2}\right] \\
&= \cos^2\theta\frac{\partial^2 f}{\partial x^2} + 2\sin\theta\cos\theta\frac{\partial^2 f}{\partial x\partial y} + \sin^2\theta\frac{\partial^2 f}{\partial y^2}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 f}{\partial \theta^2} &= \frac{\partial}{\partial \theta}\left[-r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}\right] \\
&= -r\cos\theta\frac{\partial f}{\partial x} - r\sin\theta\frac{\partial f}{\partial y} \\
&\quad -r\sin\theta\frac{\partial}{\partial \theta}\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial}{\partial \theta}\frac{\partial f}{\partial y} \\
&= -r\cos\theta\frac{\partial f}{\partial x} - r\sin\theta\frac{\partial f}{\partial y} \\
&\quad -r\sin\theta\left[-r\sin\theta\frac{\partial^2 f}{\partial x^2} + r\cos\theta\frac{\partial^2 f}{\partial x\partial y}\right] \\
&\quad +r\cos\theta\left[-r\sin\theta\frac{\partial^2 f}{\partial y\partial x} + r\cos\theta\frac{\partial^2 f}{\partial y^2}\right] \\
&= r^2\sin^2\theta\frac{\partial^2 f}{\partial x^2} - 2r^2\sin\theta\cos\theta\frac{\partial^2 f}{\partial x\partial y} + r^2\cos^2\theta\frac{\partial^2 f}{\partial y^2} \\
&\quad -r\cos\theta\frac{\partial f}{\partial x} - r\sin\theta\frac{\partial f}{\partial y}.
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\partial^2 f}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} - \frac{\cos\theta}{r}\frac{\partial f}{\partial x} - \frac{\sin\theta}{r}\frac{\partial f}{\partial y} \\
&= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} - \frac{1}{r}\frac{\partial f}{\partial r}
\end{aligned}$$

in other words

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{1}{r}\frac{\partial f}{\partial r}. \tag{6.17}$$

Therefore, under the polar coordinate system (r,θ) the Laplace operator

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{1}{r}\frac{\partial}{\partial r}. \tag{6.18}$$

6.5.2 Cylindrical coordinate system in $\mathbb{R}^3$

Let (x,y,z) be the standard coordinate system. In the cylindrical polar coordinates we keep the z-coordinate and use the polar coordinates for (x,y), that is

$$x = r\cos\theta,\ y = r\sin\theta,\ z = z$$

where $0 \le \theta < 2\pi$. The Jacobian matrix is given by

$$\begin{pmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the Jacobian $\frac{\partial(x,y,z)}{\partial(r,\theta,z)} = r$. The inverse transformation is given by

$$r = \sqrt{x^2+y^2},\ \tan\theta = \frac{y}{x},\ z = z.$$

The Laplace operator in the cylindrical coordinates is

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial z^2}. \tag{6.19}$$

6.5.3 Spherical coordinates in $\mathbb{R}^3$

Let (x,y,z) be the Cartesian coordinates for a general point $P \in \mathbb{R}^3$, $P \neq (0,0,0)$. Let ρ be the distance between P and O: $\rho = \sqrt{x^2+y^2+z^2}$, and let φ be the angle from z-axis to the position vector $\overrightarrow{OP}$, so that $z = \rho\cos\varphi$ where $0 \le \varphi \le \pi$. Change (x,y) to its polar coordinates $(r\cos\theta, r\sin\theta)$, where r is the distance from O to the perpendicular projection of P to the xy-plane, so that $r = \rho\sin\varphi$. In terms of the spherical coordinates (ρ,φ,θ) we have

$$\left\{\begin{array}{lcl} x &=& \rho\sin\varphi\cos\theta, \\ y &=& \rho\sin\varphi\sin\theta, \\ z &=& \rho\cos\varphi. \end{array}\right. \tag{6.20}$$

where $\rho \ge 0$, $0 \le \varphi \le \pi$ and $0 \le \theta \le 2\pi$. The Jacobian matrix of the transformation (6.20) can be computed directly:

$$\begin{pmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{pmatrix}.$$

Hence the Jacobian is given by

$$\begin{aligned}
\frac{\partial(x,y,z)}{\partial(\rho,\varphi,\theta)} &= \begin{vmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{vmatrix} \\
&= \rho^2\sin\varphi \begin{vmatrix} \sin\varphi\cos\theta & \cos\varphi\cos\theta & -\sin\theta \\ \sin\varphi\sin\theta & \cos\varphi\sin\theta & \cos\theta \\ \cos\varphi & -\sin\varphi & 0 \end{vmatrix} \\
&= \rho^2\sin\varphi \left(\cos\varphi \begin{vmatrix} \cos\varphi\cos\theta & -\sin\theta \\ \cos\varphi\sin\theta & \cos\theta \end{vmatrix} + \sin\varphi \begin{vmatrix} \sin\varphi\cos\theta & -\sin\theta \\ \sin\varphi\sin\theta & \cos\theta \end{vmatrix} \right) \\
&= \rho^2\sin\varphi \left(\cos^2\varphi \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} + \sin^2\varphi \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} \right) \\
&= \rho^2\sin\varphi.
\end{aligned}$$

The inverse transformation can be worked out as the following

$$\begin{cases} \rho &= \sqrt{x^2+y^2+z^2}, \\ \tan\varphi &= \frac{\sqrt{x^2+y^2}}{z}, \\ \tan\theta &= \frac{y}{x} . \end{cases}$$

Finally let us consider the Laplace operator in $\mathbb{R}^3$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

and we wish to write the Laplace operator in the spherical coordinate system. Suppose $f(x,y,z)$ has continuous derivatives up to second order. First, we use the cylindrical coordinates $x = r\cos\theta$, $y = r\sin\theta$, $z = z$. Then, according to (6.17)

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2}. \tag{6.21}$$

Next, we use the change of variables: $z = \rho\cos\varphi$ and $r = \rho\sin\varphi$. Notice that (ρ,φ) are the polar coordinates for (z,r), thus, according to (6.14)

$$\begin{cases} \frac{\partial f}{\partial \rho} &= \cos\varphi\frac{\partial f}{\partial z} + \sin\varphi\frac{\partial f}{\partial r}, \\ \frac{\partial f}{\partial \varphi} &= -\rho\sin\varphi\frac{\partial f}{\partial z} + \rho\cos\varphi\frac{\partial f}{\partial r}, \end{cases} \tag{6.22}$$

and, according to (6.17)

$$\frac{\partial^2 f}{\partial z^2} + \frac{\partial^2 f}{\partial r^2} = \frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial f}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \varphi^2}. \tag{6.23}$$

Putting (6.21) and (6.23) together to obtain

$$\begin{aligned}
\Delta f &= \frac{\partial^2 f}{\partial r^2} + \frac{\partial^2 f}{\partial z^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} \\
&= \frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial f}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \varphi^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2}. \qquad (6.24)
\end{aligned}$$

On the other hand, by solving $\frac{\partial f}{\partial r}$ from (6.22) we have

$$\frac{\partial f}{\partial r} = \sin\varphi \frac{\partial f}{\partial \rho} + \frac{\cos\varphi}{\rho}\frac{\partial f}{\partial \varphi} \tag{6.25}$$

and substituting it into (6.24) we finally obtain

$$\begin{aligned} \Delta f &= \frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial f}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \varphi^2} \\ &\quad + \frac{1}{\rho\sin\varphi}\left[\sin\varphi\frac{\partial f}{\partial \rho} + \frac{\cos\varphi}{\rho}\frac{\partial f}{\partial \varphi}\right] + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} \\ &= \frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 f}{\partial \varphi^2} + \frac{1}{\rho^2\cos^2\varphi}\frac{\partial^2 f}{\partial \theta^2} + \frac{2}{\rho}\frac{\partial f}{\partial \rho} + \frac{\cot\varphi}{\rho^2}\frac{\partial f}{\partial \varphi}. \end{aligned}$$

That is, under the spherical coordinate system the Laplace operator in $\mathbb{R}^3$ can be written as

$$\Delta = \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2}{\partial \varphi^2} + \frac{1}{\rho^2\cos^2\varphi}\frac{\partial^2}{\partial \theta^2} + \frac{2}{\rho}\frac{\partial}{\partial \rho} + \frac{\cot\varphi}{\rho^2}\frac{\partial}{\partial \varphi}. \tag{6.26}$$

6.6 Some simple partial differential equations

An equation involving several variables, functions and their partial derivatives is called a partial differential equation (abbreviated as PDE or PDEs for simplicity). Some important examples of PDES may be listed as the following

- Laplace's equation (in steady-state temperature, electrostatic potential, fluid flow etc.)

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0\ .$$

  In the cylindrical coordinates r, θ, z given by $x = r\cos\theta$, $y = r\sin\theta$, $z = z$, the equation is

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

  If we use the spherical coordinates ρ, φ, θ defined by $x = \rho\sin\varphi\cos\theta$, $y = \rho\sin\varphi\sin\theta$ and $z = \rho\cos\varphi$, then Laplace's equation becomes

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{2}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \varphi^2} + \frac{\cot\varphi}{\rho^2}\frac{\partial u}{\partial \varphi} + \frac{\csc\varphi}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} = 0.$$

- Heat equation (modelling the distribution of temperature, the diffusion of heat)

$$\frac{\partial u}{\partial t} = \nu^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right)$$

  in which ν^2 is a constant called *termal diffusivity.*

- Schrödinger's wave equation in quantum physics

$$-\frac{\hbar}{2m}\Delta\psi + V(x)\psi = E\psi$$

where $\hbar$ is the Planck's constant, m is the mass, V is the potential energy and E is the energy. ψ represents the wave function. For further reading, see for example, A. P. French and E.F. Taylor: An Introduction to Quantum Physics.

- Black-Scholes partial differential equation in finance

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

where S represents the current value of an asset, r is the interest rate for a riskless account, σ^2 represents the volatility. For further reading, see for example, P. Wilmott, S. Howison and J. Dewynne: The mathematics of financial derivatives – a student introduction.

- Maxwell's equations for electromagnetic fields

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0; \; \nabla \cdot \mathbf{B} = 0; \\ \frac{\partial}{\partial t}\mathbf{B} &= -\nabla \times \mathbf{E}; \; \frac{1}{c^2}\frac{\partial}{\partial t}\mathbf{E} = \nabla \times \mathbf{B} - \mu_0 \mathbf{J} \end{aligned}$$

where $\mathbf{E}$ is the electric field strength, $\mathbf{B}$ the magnetic field strength, $\mathbf{J}$ the volume current density, c is the speed of light, and μ_0 the permeability of vacuum. For more details, see for example, P. Lorrain, D. Corson and F. Lorrain: Electromagnetic fields and waves.

- Navier-Stokes equations for incompressible flow of fluid, which is a system of PDEs for the velocity vector field $\mathbf{u} = (u^1, u^2, u^3)$ and the pressure p:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = \nu \Delta \mathbf{u} - \nabla p, \;\; \nabla \cdot \mathbf{u} = 0,$$

where the contraction $\mathbf{u} \cdot \nabla \mathbf{u}$ is the directional derivatives of the velocity $\mathbf{u}$ in the direction $\mathbf{u}$. That is

$$\mathbf{u} \cdot \nabla \mathbf{u} = (\mathbf{u} \cdot \nabla u^1, \mathbf{u} \cdot \nabla u^2, \mathbf{u} \cdot \nabla u^3)$$

where

$$\mathbf{u} \cdot \nabla u^i = D_{\mathbf{u}} u^i = u^1 \frac{\partial u^i}{\partial x} + u^2 \frac{\partial u^i}{\partial y} + u^3 \frac{\partial u^i}{\partial z}$$

for $i = 1, 2, 3$. In terms of the components u^1, u^2, u^3 of the velocity $\mathbf{u}$, the Navier-Stokes equations are

$$\begin{aligned} \frac{\partial u^1}{\partial t} + u^1 \frac{\partial u^1}{\partial x} + u^2 \frac{\partial u^1}{\partial y} + u^3 \frac{\partial u^1}{\partial z} &= \nu \Delta u^1 - \frac{\partial p}{\partial x}, \\ \frac{\partial u^2}{\partial t} + u^1 \frac{\partial u^2}{\partial x} + u^2 \frac{\partial u^2}{\partial y} + u^3 \frac{\partial u^2}{\partial z} &= \nu \Delta u^2 - \frac{\partial p}{\partial y}, \\ \frac{\partial u^3}{\partial t} + u^1 \frac{\partial u^3}{\partial x} + u^2 \frac{\partial u^3}{\partial y} + u^3 \frac{\partial u^3}{\partial z} &= \nu \Delta u^3 - \frac{\partial p}{\partial z}, \\ \frac{\partial u^1}{\partial x} + \frac{\partial u^2}{\partial y} + \frac{\partial u^3}{\partial z} &= 0. \end{aligned}$$

Of course there are many more PDEs which appear in science. In this course we will study none of them, and instead we choose to study some very simple PDEs as examples for which we even can find closed forms (i.e. explicit formulate) for their solutions.

6.6.1 First order linear PDE (in two variables)

Let us consider the following type of partial differential equations (in two variables)

$$P(x,y)\frac{\partial z}{\partial x}+Q(x,y)\frac{\partial z}{\partial y}=0 \tag{6.27}$$

where at least one of P and Q does not vanish. We wish to find some explicit solutions to (6.27). The idea is to consider the first order ODE

$$\frac{dx}{P(x,y)}=\frac{dy}{Q(x,y)} \ . \tag{6.28}$$

That is,

$$\frac{dy}{dx}=\frac{Q(x,y)}{P(x,y)}.$$

Suppose the general solution to (6.28) is given implicitly by $\psi(x,y)=C$ where C is a constant, then the general solution to the PDE (6.28) is given by

$$z=\Phi(\psi(x,y))$$

where Φ is an arbitrary differentiable function. In fact, if $\psi(x,y)=C$ is the solution to (6.28) so that, by differentiating in x we obtain

$$\frac{\partial \psi}{\partial x}+\frac{\partial \psi}{\partial y}\frac{dy}{dx}=\frac{d}{dx}C=0.$$

Substituting $\frac{dy}{dx}=\frac{Q}{P}$ into the equation to obtain

$$\frac{\partial \psi}{\partial x}+\frac{\partial \psi}{\partial y}\frac{Q}{P}=0 \ .$$

On the other hand, by the chain rule

$$\frac{\partial z}{\partial x}=\Phi'\frac{\partial \psi}{\partial x}, \ \frac{\partial z}{\partial y}=\Phi'\frac{\partial \psi}{\partial y}$$

so that

$$P(x,y)\frac{\partial z}{\partial x}+Q(x,y)\frac{\partial z}{\partial y}=\Phi'\left(P\frac{\partial \psi}{\partial x}+Q\frac{\partial \psi}{\partial y}\right)=0,$$

which shows that $z=\Phi(\psi(x,y))$ is a solution to the PDE (6.27).

Example 6.14 *Find the solution to the following PDE*

$$x\frac{\partial z}{\partial y}-y\frac{\partial z}{\partial x}=0$$

such that when $x=0$ *then* $z=y^2$.

First solve the first order ODE

$$\frac{dx}{-y} = \frac{dy}{x}$$

which is separable, and has the general solution given by $x^2 + y^2 = C$. Thus the solution to the PDE is $z = \Phi(x^2 + y^2)$, where Φ is differentiable. We want a solution such that when $x = 0$, $z = y^2$ so that $\Phi(y^2) = y^2$ for any y, hence $\Phi(t) = t$ for $t \geq 0$, which yields that $z = x^2 + y^2$.

6.6.2 First order quasi linear PDEs (in two variables)

Similarly, in order to solve the following type of first order linear PDEs

$$P(x,y,z)\frac{\partial z}{\partial x} + Q(x,y,z)\frac{\partial z}{\partial y} = R(x,y,z) \tag{6.29}$$

we may attempt to solve the following ODEs

$$\frac{dx}{P(x,y,z)} = \frac{dy}{Q(x,y,z)} = \frac{dz}{R(x,y,z)}$$

which contains three ODES, but in general only two of them are independent. Choose a pair of them, solve them and obtain two solutions which may be given implicitly by

$$\psi_1(x,y,z) = C_1 \text{ and } \psi_2(x,y,z) = C_2.$$

Then the general solution to the PDE (6.29) is given implicitly by

$$\Phi(\psi_1(x,y,z), \psi_2(x,y,z)) = 0$$

where Φ is an arbitrary function with continuous partial derivatives.

Example 6.15 *Find the solution to the following PDE*

$$x\frac{\partial z}{\partial x} + (y + x^2)\frac{\partial z}{\partial y} = z$$

which satisfies the condition that when $x = 2$ then $z = y - 4$.

Set up the auxiliary ODEs

$$\frac{dx}{x} = \frac{dy}{y + x^2} = \frac{dz}{z}.$$

From which choose two of them, for example

$$\frac{dx}{x} = \frac{dy}{y + x^2}, \; \frac{dx}{x} = \frac{dz}{z}.$$

The first equation may be written as

$$\frac{dy}{dx} - \frac{1}{x}y = x$$

which is a first order linear equation. $\frac{1}{x}$ is an integrating factor, multiplying $\frac{1}{x}$ to obtain

$$\frac{1}{x}\frac{dy}{dx} - \frac{1}{x^2}y = \frac{d}{dx}\left(\frac{y}{x}\right) = 1,$$

integrating the equation to obtain $\frac{y}{x} = x + C_1$, i.e.

$$\frac{y - x^2}{x} = C_1$$

is the general solution to the first ODE. The second ODE has a general solution

$$\frac{z}{x} = C_2.$$

Therefore the general solution to the PDE is

$$\Phi(\frac{y - x^2}{x}, \frac{z}{x}) = 0,$$

or by solving $\frac{z}{x}$, the solution may be written as

$$z = xf\left(\frac{y - x^2}{x}\right)$$

where f is an arbitrary differentiable function. If $x = 2$ then $z = y - 4$ so that

$$y - 4 = 2f\left(\frac{y - 4}{2}\right).$$

Thus $t = 2f(t/2)$ (after substitution $t = y - 4$) so that $f(t) = t$. Hence the solution to the initial problem is given by $z = y - x^2$.

6.6.3 Method of separation of variables

Some PDEs (in two variables) have special type of solutions which are separable, i.e. in a product form $u(x,t) = g(x)h(t)$, for which we may attempt to make substitution: $u(x,t) = g(x)h(t)$ to reduce the PDE to ordinary differential equations for g and h.

Example 6.16 *(The heat equation) Consider the one-dimensional heat equation:*

$$\frac{\partial u(x,t)}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 u(x,t)}{\partial x^2}$$

where $\sigma > 0$ is a constant.

By an inspection, we can see that the Gaussian probability function

$$u(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{x^2}{2\sigma^2 t}} \quad \text{for } t > 0$$

is a positive solution to the heat equation for $t > 0$.

Let us search for solutions $u(x,t)$ which are separable. To this end, make substitution $u(x,t) = g(x)h(t)$. Since

$$\frac{\partial u(x,t)}{\partial t} = g(x)h'(t), \text{ and } \frac{\partial^2 u(x,t)}{\partial x^2} = g''(x)h(t)$$

thus, the heat equation becomes

$$g(x)h'(t) = \frac{\sigma^2}{2} g''(x)h(t).$$

Separate the variables to obtain

$$\frac{h'(t)}{h(t)} = \frac{\sigma^2}{2} \frac{g''(x)}{g(x)}.$$

Since $\frac{h'(t)}{h(t)}$ only depends on t, while the equation implies that it only depends on x, therefore $\frac{h'(t)}{h(t)}$ must be a constant function. Similarly, $\frac{g''(x)}{g(x)}$ is a constant independent of x or t. Therefore we must have

$$\frac{h'(t)}{h(t)} = \frac{\sigma^2}{2} \frac{g''(x)}{g(x)} = \lambda$$

where λ is a constant. The heat equation is thus transformed to a system of second order linear ODEs

$$h'(t) = \lambda h(t), \;\; g''(x) = \frac{2\lambda}{\sigma^2} g(x).$$

The first ODE has a general solution $h(t) = C_1 e^{\lambda t}$ and the second ODE has a general solution

1. If $\lambda > 0$, then
$$g(x) = C_2 e^{-\sqrt{\frac{2\lambda}{\sigma^2}}x} + C_3 e^{\sqrt{\frac{2\lambda}{\sigma^2}}x}.$$

2. If $\lambda < 0$, then
$$g(x) = C_2 \cos\left(\sqrt{-\frac{2\lambda}{\sigma^2}}x\right) + C_3 \sin\left(\sqrt{-\frac{2\lambda}{\sigma^2}}x\right).$$

3. If $\lambda = 0$, then
$$g(x) = C_2 + C_3 x.$$

7 Gradient vectors, normal vectors to surfaces

In this part we consider curves and surfaces in $\mathbb{R}^3$. For simplicity, let us declare that all functions we will encounter in this part are defined on open subsets (unless otherwise specified), and have continuous partial derivatives.

The graph of a function $y = f(x)$ defined on (a, b) is a curve in the plane $\mathbb{R}^2$. The derivative $f'(x_0)$ measures the slope of the line tangent to the graph at $(x_0, f(x_0))$: $\tan\alpha = f'(x_0)$ where α is the angle from the x-axis to the tangent line at $(x_0, f(x_0))$. The equation for the tangent line at $(x_0, f(x_0))$ is a linear equation

$$y - f(x_0) = f'(x_0)(x - x_0).$$

The graph of $y = f(x)$ has a natural *parameterization*: we may write the coordinates (x, y) on the graph as the following

$$x = t, \;\; y = f(t) \tag{7.1}$$

and consider $t \in (a, b)$ as a parameter. The mapping $t \to (t, f(t))$ is called a *parameterized curve* in the plane. Similarly, the tangent line at $(x_0, f(x_0))$ has a natural parameterization, namely given as

$$x = t,\ y = f(x_0) + f'(x_0)(t - x_0),$$

where (x, y) is a general point lying on the tangent line. In terms of vector notations, it can be written as

$$(x - x_0, y - f(x_0)) = (1, f'(x_0))(t - x_0)$$

i.e. $(x - x_0, y - f(x_0))$ is parallel to the vector $(1, f'(x_0))$ which is called the *tangent vector* of the parameterized curve. Note that the first coordinate 1 appears as $\frac{dx}{dt} = 1$, so the tangent vector at $(x(t), y(t))$ can be written as $(x'(t), y'(t))$, where $x(t) = t$ and $y(t) = f(t)$ for the graph of $y = f(x)$.

Generalized this notion to give the concept of parameterized curves in the plane $\mathbb{R}^2$. That is, a parameterized curve γ in the plane $\mathbb{R}^2$ is a mapping $t \to (x(t), y(t))$, i.e.

$$x = x(t),\ y = y(t),$$

where $t \in (a, b)$ (some interval) is a parameter. Since $\frac{dy}{dx} = \frac{y'(t)}{x'(t)}$, so the tangent line to the curve γ at a point $(x(t_0), y(t_0))$ has a parameterization

$$\begin{cases} x = x(t_0) + x'(t_0)(t - t_0); \\ y = y(t_0) + y'(t_0)(t - t_0) \end{cases}$$

which represents the line passing through $(x(t_0), y(t_0))$ with slope $(x'(t_0), y'(t_0))$.

A curve in $\mathbb{R}^2$ can also be described implicitly by an equation such as $F(x, y) = 0$. You should be familiar with the standard quadratic curves such as circles, ellipses, parabolas and hyperbolas (for a revision you may refer to Richard Earl's notes).

Example 7.1 *Consider an ellipse defined implicitly by the equation*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

which has a parameterization defined by

$$x = a\cos t, \quad y = b\sin t$$

where $0 \le t < 2\pi$*. A tangent vector at* $(a\cos t,\ \ b\sin t)$ *is thus given as* $(-a\sin t, b\cos t)$*.*

A parameterized curve γ in the space $\mathbb{R}^3$ may be described by a vector valued function of one variable t, i.e. a mapping $t \to \gamma(t)$ where

$$\gamma(t) = (x(t), y(t), z(t)), \quad t \in (a, b). \tag{7.2}$$

The *tangent vector* to the curve at $\gamma(t_0)$ is the vector $(x'(t_0), y'(t_0), z'(t_0))$ and the line tangent to the curve at $(x(t_0), y(t_0), z(t_0))$ is given by the equation

$$r = \gamma(t_0) + \gamma'(t_0)(t - t_0).$$

that is

$$\begin{cases} x = x(t_0) + x'(t_0)(t - t_0), \\ y = y(t_0) + y'(t_0)(t - t_0), \\ z = z(t_0) + z'(t_0)(t - t_0). \end{cases}$$

7.1 Normal vectors, tangent planes

Let us consider smooth surfaces in $\mathbb{R}^3$. As for the case of curves, the graph of a function $z = f(x, y)$ on a domain U is considered as a *parameterized surface*

$$(x, y) \to (x, y, f(x, y)), \qquad (x, y) \in U. \tag{7.3}$$

By relabel the variables, the graph of $z = f(x, y)$ is a parameterized surface defined by the mapping $(u, v) \to (u, v, f(u, v))$ where (u, v) as two parameters. In general, a mapping

$$(u, v) \to (x(u, v), y(u, v), z(u, v)) \tag{7.4}$$

where (u, v) runs through an open subset $U \subset \mathbb{R}^2$ is called a parameterized surface in the space $\mathbb{R}^3$. The mapping or the parameterized surface is often written as

$$x = x(u, v),\ y = y(u, v),\ \ z = z(u, v).$$

When (u, v) runs through a subset U, then $(x(u, v), y(u, v), z(u, v))$ draws out a surface in the space the image of U under the mapping (7.4).

A surface S may be described by an equation

$$F(x, y, z) = 0, \tag{7.5}$$

where, in order to avoid technical difficulty, we assume that $\nabla F \neq 0$. For example, a sphere: $x^2 + y^2 + z^2 = R^2$ which has a parameterized representation in terms of spherical coordinates (φ, θ) [Notice that the equation of the sphere in spherical coordinates takes a simple form: $\rho = R$]

$$x = R \sin\varphi \cos\theta, \quad y = R \sin\varphi \sin\theta, \ \ z = R\cos\varphi,$$

where $\varphi \in [0, \pi]$ and $\theta \in [0, 2\pi)$ are two parameters.

The graph of a function $z = f(x, y)$ is a surface which is a parameterized surface, but also can be described by the equation

$$z - f(x, y) = 0.$$

Let us now define the concept of the *tangent plane* at a point on the surface. Let $P = (x_0, y_0, z_0) \in S$ the surface defined by the equation (7.5) and $\gamma(t) = (x(t), y(t), z(t))$ be any parameterized curve on the surface S passing though the point P, say $\gamma(0) = (x_0, y_0, z_0) = P$. Then

$$F(x(t), y(t), z(t)) = 0 \qquad \forall t$$

so, by differentiating in t at $t = 0$, employing the chain rule, we obtain

$$F_x(x_0, y_0, z_0)x'(0) + F_y(x_0, y_0, z_0)y'(0) + F_z(x_0, y_0, z_0)z'(0) = 0. \tag{7.6}$$

Recall that $\nabla F = (F_x, F_y, F_z)$ is the gradient vector field of F, so we may rewrite (7.10) as

$$\nabla F(x_0, y_0, z_0) \cdot \gamma'(0) = 0 \tag{7.7}$$

which says the tangent vector $\gamma'(0)$ is perpendicular to the gradient of F. Since γ can be any curve on the surface S, thus $\gamma'(0)$ can be any vector tangent to the surface S at P, so (7.7) means that

any vector tangent to the surface S at P is perpendicular to the gradient vector $\nabla F(x_0, y_0, z_0)$, and therefore all tangent vectors to S at the point P lies on the plane passing through P and perpendicular to $\nabla F(x_0, y_0, z_0)$, which is called the *tangent plane* to S at P. We therefore call $\nabla F(x_0, y_0, z_0)$ a *normal vector* to the surface S at P.

Suppose that (x, y, z) belongs to the tangent plane at P, so that $(x - x_0, y - y_0, z - z_0)$ lies on the tangent plane, so it must be perpendicular to the normal vector $\nabla F(x_0, y_0, z_0)$, thus

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0 \tag{7.8}$$

which is the equation for the tangent plane to S at $P = (x_0, y_0, z_0)$.

If the surface S is the graph of a function $z = f(x, y)$, so that we may take $F(x, y, z) = z - f(x, y)$ thus a normal vector at $(x, y, f(x, y))$ is the gradient vector of F which is $(-f_x, -f_y, 1)$. Therefore an equation for the tangent plane to the graph of $z = f(x, y)$ at $(x_0, y_0, f(x_0, y_0))$ is given by

$$-f_x(x_0, y_0)(x - x_0) - f_y(x_0, y_0)(x - x_0) + z - z_0 = 0 \tag{7.9}$$

which we have already seen before.

Consider on the other hand a parameterized surface S: which is described by a vector valued function of two parameters (u, v):

$$x = x(u, v),\ y = y(u, v) \text{ and } z = z(u, v) \tag{7.10}$$

where (u, v) runs through an open subset $D \subseteq \mathbb{R}^2$. Let $(u_0, v_0) \in D$ and

$$P = (x_0, y_0, z_0) = (x(u_0, v_0), y(u_0, v_0), z(u_0, v_0))$$

is a point on the surface S. Consider the parameterized curve γ_1 defined by

$$\gamma_1(u) = (x(u, v_0), y(u, v_0), z(u, v_0))$$

and the parameterized curve γ_2 defined by

$$\gamma_2(v) = (x(u_0, v), y(u_0, v), z(u_0, v))$$

where u (resp. v) is considered as a parameter. Then both curves γ_1 and γ_2 lie on the surface and pass through P, and the tangent vectors $\gamma_1'(u_0)$ and $\gamma_2'(v_0)$ are two tangent vectors to the parameterized surface S at P, thus, by definition $\gamma_1'(u_0) \times \gamma_2'(v_0)$ (cross product of two vectors $\gamma_1'(u_0) \times \gamma_2'(v_0)$) is a vector perpendicular to the both vectors $\gamma_1'(u_0)$ and $\gamma_2'(v_0)$ [Geometry I, Prelims] and therefore $\gamma_1'(u_0) \times \gamma_2'(v_0)$ is a normal vector to the surface S. On the other hand, by definition of partial derivatives

$$\gamma_1' = \frac{\partial \gamma}{\partial u} = \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u} \right),\ \gamma_2' = \frac{\partial \gamma}{\partial v} = \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v} \right)$$

where $\gamma(u, v) = (x(u, v), y(u, v), z(u, v))$. Thus $\frac{\partial \gamma}{\partial u} \times \frac{\partial \gamma}{\partial v}$ is a normal vector to the parameterized surface S, which is given by according to the definition of the cross product

$$\frac{\partial \gamma}{\partial u} \times \frac{\partial \gamma}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{vmatrix}.$$

The tangent plane to S at $\gamma(u_0, v_0)$ has an equation

$$\frac{\partial\gamma(u_0,v_0)}{\partial u} \times \frac{\partial\gamma(u_0,v_0)}{\partial v} \cdot (\mathbf{r} - \gamma(u_0,v_0)) = 0 \tag{7.11}$$

where $\mathbf{r} = (x, y, z)$ is the position vector on for a general point in the tangent plane. In terms of the Cartesian coordinates, (7.11) can be written, by working out the dot product, as

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{vmatrix} = 0 \tag{7.12}$$

where the partial derivatives are evaluated at (u_0, v_0), and $(x_0, y_0, z_0) = \gamma(u_0, v_0)$.

Example 7.2 *The sphere with radius $R > 0$ may be described implicitly by the equation*

$$x^2 + y^2 + z^2 = R^2,$$

so a normal vector to the tangent plane at (x_0, y_0, z_0) is $\nabla f(x_0, y_0, z_0) = 2(x_0, y_0, z_0)$ which has the same direction as the coordinate vector, and the tangent plane has an equation

$$x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0.$$

Since the point (x_0, y_0, z_0) lies on the sphere so that the equation can be simplified as

$$x_0 x + y_0 y + z_0 z = R^2.$$

The sphere may be parameterized via the spherical coordinates which are is given as the parameterized surface

$$x = R\sin\varphi\cos\theta,\ y = R\sin\varphi\sin\theta,\ z = R\cos\varphi$$

where $0 \le \varphi \le \pi$ and $0 \le \theta < 2\pi$, hence the tangent plane at a point (x_0, y_0, z_0) has an equation

$$\begin{aligned} \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi} & \frac{\partial z}{\partial \varphi} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta} \end{vmatrix} &= \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ R\cos\varphi\cos\theta & R\cos\varphi\sin\theta & -R\sin\varphi \\ -R\sin\varphi\sin\theta & R\sin\varphi\cos\theta & 0 \end{vmatrix} \\ &= R^2 \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ \cos\varphi\cos\theta & \cos\varphi\sin\theta & -\sin\varphi \\ -\sin\varphi\sin\theta & \sin\varphi\cos\theta & 0 \end{vmatrix} = 0 \end{aligned}$$

which may be simplified as the following

$$\sin\varphi\cos\theta\,(x - x_0) + \sin\varphi\sin\theta\,(y - y_0) + \cos\varphi\,(z - z_0) = 0.$$

7.2 Directional derivatives

Suppose that $F(x, y, z)$ has continuous partial derivatives, then its gradient vector by definition is $\nabla F = (F_x, F_y, F_z)$. Suppose $\gamma(t) = (x(t), y(t), z(t))$ (where $t \in (a, b)$) is a parameterized curve with continuous derivatives, then

$$f(t) = F \circ \gamma(t) = F(x(t), y(t), z(t))$$

is differentiable, and, by chain rule,

$$\begin{aligned}\frac{df}{dt} &= F_x x'(t) + F_y y'(t) + F_z z'(t) \\ &= \nabla F(\gamma(t)) \cdot \gamma'(t).\end{aligned}$$

In particular, if $\mathbf{v} = (v_1, v_2, v_3)$ is a no-zero vector, and take

$$\gamma(t) = (x_0 + v_1 t, y_0 + v_2 t, z_0 + v_3 t)$$

the line passing through (x_0, y_0, z_0), then the derivative

$$\begin{aligned}\frac{d}{dt} F \circ \gamma(0) &= \nabla F \cdot \gamma'(0) \\ &= \nabla F(x_0, y_0, z_0) \cdot \mathbf{v} \\ &= v_1 F_x + v_2 F_y + v_3 F_z\end{aligned}$$

is called the *directional derivative* of F in $\mathbf{v}$, denoted by $D_{\mathbf{v}} F$, hence

$$D_{\mathbf{v}} F = \nabla F \cdot \mathbf{v}.$$

By definition

$$D_{\mathbf{v}} F(x_0, y_0, z_0) = \lim_{t \to 0} \frac{F(x_0 + v_1 t, y_0 + v_2 t, z_0 + v_3 t) - F(x_0, y_0, z_0)}{t}.$$

The previous discussion can be stated as the following

Proposition 7.3 *Suppose that $F(x, y, z)$ is a function on an open subset $U \subset \mathbb{R}^3$ with continuous partial derivatives, and $\gamma(t)$ is a parameterized curve in U with tangent vector $\mathbf{v}$ at $\gamma(0)$, i.e. $\gamma'(0) = \mathbf{v}$, then*

$$\frac{d}{dt} F \circ \gamma(0) = D_{\mathbf{v}} F(\gamma(0)) = \nabla F(\gamma(0)) \cdot \mathbf{v}. \tag{7.13}$$

8 Taylor's theorem

Suppose that $f(x)$ is a function defined on $[a, b]$ with derivatives of any order. For a given natural number n we search for a polynomial in $(x - a)$ of degree n

$$p_n(x) = a_0 + a_1(x - a) + \cdots + a_n(x - a)^n$$

so that $f(x)$ agrees with $p_n(x)$ up to nth order derivatives at a, that is $f^{(k)}(a) = p^{(k)}(a)$ for $k = 0, 1, \cdots, n$. Since $p^{(k)}(a) = k!a_k$ for $k = 0, \cdots, n$ we obtain $a_k = \frac{1}{k!}f^{(k)}(a)$ and therefore

$$p_n(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{8.1}$$

which is called *Taylor's expansion* (of order n) for f at the point a. We have the following theorem which will be proved in Prelims Analysis II in Hilary term.

Theorem 8.1 *(Taylor's theorem for one variable function) Suppose $f(x)$ has derivatives at a up to nth order, then*

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + o((x-a)^n)) \tag{8.2}$$

as $x \to a$ [the right-hand side is called Taylor's expansion of f at a with Peano's remainder]. That is

$$\lim_{x \to a} \frac{f(x) - p_n(x)}{(x-a)^n} = 0.$$

Taylor's theorem says the Taylor expansion of nth order is a good approximation of f near a up to $(x-a)^n$.

We can have better estimate for the difference $f(x) - p_n(x)$ if f has derivatives on $[a, b]$ up to $(n+1)$th order. Namely we have

Theorem 8.2 (Taylor's Theorem) *Suppose $f(x)$ has derivatives on $[a, b]$ up to $(n+1)$th order, then for any $x \in (a, b]$ there is $\xi \in (a, x)$ such that*

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \tag{8.3}$$

In particular, if f has derivatives of any order, and if

$$\frac{M_n}{n!}(b-a)^n \to 0 \qquad \text{as } n \to \infty$$

where $M_n = \sup_{[a,b]} |f^{(n)}(x)|$, then

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots \qquad \forall x \in [a, b].$$

For example, we can easily see that

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots \quad \forall x \in (-\infty, \infty).$$

Let us now consider a function $f(x, y)$ of two variables defined on an open subset U. Suppose $(x_0, y_0) \in U$. We search for a Taylor type expansion of $f(x, y)$ near (x_0, y_0). Let us assume that

f has continuous partial derivatives up to order n. Let $(x,y) \in U$ close to (x_0,y_0) so that the line segment

$$\begin{aligned}\gamma(t) &= (1-t)(x_0,y_0)+t(x,y)\\ &= (x_0,y_0)+t(x-x_0,y-y_0)\end{aligned}$$

(where $t \in [0,1]$) between (x_0,y_0) and (x,y) lies in U. Consider one variable function

$$g(t) = f\circ\gamma(t) \qquad t\in[0,1].$$

Then g has derivatives on $[0,1]$ up to order n, so we can apply Taylor's Theorem at $a=0$. To simplify our computations below we introduce vector notation $\mathbf{v} = (x-x_0, y-y_0)$. We want to calculate $g^{(k)}(0)$ for $k = 0,1, \cdots$. Clearly $g(0) = f(x_0,y_0)$ and

$$g'(0) = \nabla f(x_0,y_0)\cdot\mathbf{v}$$

as we have seen in the previous sections. In general

$$\begin{aligned}g'(t) &= \nabla f(\gamma(t))\cdot\gamma'(t) = \nabla f(\gamma(t))\cdot\mathbf{v}\\ &= f_x(\gamma(t))(x-x_0)+f_y(\gamma(t))(y-y_0)\end{aligned}$$

so that, by differentiating in t again to obtain

$$\begin{aligned}g''(t) &= (x-x_0)\nabla f_x(\gamma(t))\cdot\mathbf{v}+(y-y_0)\nabla f_y(\gamma(t))\cdot\mathbf{v}\\ &= (x-x_0)\left(f_{xx}(\gamma(t))(x-x_0)+f_{xy}(\gamma(t))(y-y_0)\right)\\ &\quad +(y-y_0)\left(f_{yx}(\gamma(t))(x-x_0)+f_{yy}(\gamma(t))(y-y_0)\right)\\ &= f_{xx}(\gamma(t))(x-x_0)^2+f_{xy}(\gamma(t))(x-x_0)(y-y_0)\\ &\quad +f_{xy}(\gamma(t))(y-y_0)(x-x_0)+f_{yy}(\gamma(t))(y-y_0)^2,\end{aligned}$$

and from which we can see the pattern for kth derivative, namely

$$g^{(k)}(t) = \sum_{\substack{i+j=k\\ i,j\geq 0}} \begin{pmatrix} k\\ i\end{pmatrix} \frac{\partial^k f(\gamma(t))}{\partial x^i\partial y^j}(x-x_0)^i(y-y_0)^j$$

and therefore

$$g^{(k)}(0) = \sum_{\substack{i+j=k\\ i,j\geq 0}} \frac{k!}{i!j!}\frac{\partial^k f(x_0,y_0)}{\partial x^i\partial y^j}(x-x_0)^i(y-y_0)^j. \tag{8.4}$$

According to Taylor's theorem for one variable function applying to g at $a=0$:

$$\begin{aligned}f(x,y) &= f(x_0,y_0)+\sum_{k=1}^{n}\frac{1}{k!}\sum_{\substack{i+j=k\\ i,j\geq 0}} \begin{pmatrix} k\\ j\end{pmatrix} \frac{\partial^k f(x_0,y_0)}{\partial x^i\partial y^j}(x-x_0)^i(y-y_0)^j\\ &\quad +o(|(x-x_0,y-y_0)|)\end{aligned} \tag{8.5}$$

as $(x,y)\to(x_0,y_0)$. If f has partial derivatives on U up to $(n+1)$th order, and the segment between (x_0,y_0) and (x,y) lies in U, then there is $\theta\in(0,1)$ (depending on n, $(x_0,y_0),(x,y)$ and the function f) such that

$$\begin{aligned} f(x,y) &= f(x_0,y_0)+\sum_{k=1}^{n}\sum_{\substack{i+j=k\\ i,j\geq 0}}\frac{1}{i!j!}\frac{\partial^k f(x_0,y_0)}{\partial x^i\partial y^j}(x-x_0)^i(y-y_0)^j \\ &\quad +\sum_{\substack{i+j=n+1\\ i,j\geq 0}}\frac{1}{i!j!}\frac{\partial^{n+1}f(\xi)}{\partial x^i\partial y^j}(x-x_0)^i(y-y_0)^j \end{aligned} \tag{8.6}$$

where

$$\xi=\theta(x_0,y_0)+(1-\theta)(x,y).$$

The right-hand side of (8.6) is called the Taylor expansion of two variable function $f(x,y)$ at (x_0,y_0). To memorize this formula, you should compare it with the Binomial expansion

$$(a+b)^k=\sum_{\substack{i+j=k\\ i,j\geq 0}}\frac{k!}{i!j!}a^ib^j$$

which corresponds to the kth derivative term in the Taylor expansion. But notice that the combination numbers in binomial expansion are $\frac{k!}{i!j!}$ but in Taylor's expansion, they turn out to be $\frac{1}{i!j!}$.

It is particularly interesting for $n=1$. For simplicity, suppose $U=B_R(x_0,y_0)$ is an open disk centered at (x_0,y_0) with radius $R>0$. Suppose all first and second partial derivatives are continuous on U. Then for any $(x,y)\in U$ there is $\xi\in U$ such that

$$\begin{aligned} f(x,y) &= f(x_0,y_0)+\nabla f(x_0,y_0)\cdot(x-x_0,y-y_0) \\ &\quad +\frac{1}{2}\left[f_{xx}(\xi)(x-x_0)^2+2f_{xy}(\xi)(x-x_0)(y-y_0)+f_{yy}(\xi)(y-y_0)^2\right]. \end{aligned} \tag{8.7}$$

The remainder in (8.7) appears as a quadratic form in $x-x_0$ and $y-y_0$ with coefficients the second partial derivatives, which can be written in terms of matrix multiplication as

$$(x-x_0,y-y_0)\begin{pmatrix} f_{xx} & f_{xy}\\ f_{xy} & f_{yy}\end{pmatrix}\begin{pmatrix} x-x_0\\ y-y_0\end{pmatrix}.$$

The square matrix

$$\begin{pmatrix} f_{xx} & f_{xy}\\ f_{xy} & f_{yy}\end{pmatrix}$$

is called the hessian matrix of f, denoted by $D^2 f$.

Similarly we may write down the Taylor expansion for a several variable function. Suppose $f(x_1,\cdots,x_k)$ is defined on a ball $B_r(a)$ centered at $\mathbf{a}=(a_1,\cdots,a_k)$ with radius $r>0$, with continuouse partial derivatives of any order. Then, for any $\mathbf{x}=(x_1,\cdots,x_k)$ and $n=1,2,\ldots$, we

have

$$
\begin{aligned}
f(\mathbf{x}) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \cdots \\
&\quad + \frac{1}{n!} \sum_{\substack{i_1 + \cdots + i_k = n \\ 0 \le i_1, \cdots, i_k \le n}} \left(\begin{array}{c} n \\ i_1, \cdots, i_k \end{array} \right) \frac{\partial^n f(\mathbf{a})}{\partial x_1^{i_1} \cdots \partial x_k^{i_k}} (x_1 - a_1)^{i_1} \cdots (x_k - a_k)^{i_k} \\
&\quad + o(|x - a|^{n+1})
\end{aligned}
$$

as $\mathbf{x} \to \mathbf{a}$, where

$$
\left(\begin{array}{c} n \\ i_1, \cdots, i_k \end{array} \right) = \frac{n!}{i_1! \cdots i_k!}.
$$

9 Critical points

In this part we apply Taylor's theorem to the study of multi-variable functions near *critical points*. For simplicity, we concentrate on two variable functions, though the techniques we are going to develop apply to several variable functions with necessary modifications.

First of all we introduce the notions of local extrema. Let $f(x, y)$ be a function defined on a subset $A \subset \mathbb{R}^2$. Then a point $(x_0, y_0) \in A$ is a *local maximum* (resp. *local minimum*) of f , if there is an open ball $B_r(x_0, y_0) \subset A$ for some $r > 0$ such that

$$
f(x, y) \le f(x_0, y_0) \quad \forall (x, y) \in B_r(x_0, y_0) \tag{9.1}
$$

(resp.

$$
f(x, y) \ge f(x_0, y_0) \quad \forall (x, y) \in B_r(x_0, y_0)). \tag{9.2}
$$

On the other hand, we say $(x_0, y_0) \in A$ is a (global) maximum (resp. (global) minimum) if $f(x, y) \le f(x_0, y_0)$ (resp. $f(x, y) \ge f(x_0, y_0)$) for every $(x, y) \in A$. We should note that a global maximum (or a global minimum) for a function is not necessary a local one, for example consider the function $f(x, y) = x^2 + y^2$ defined on $A = \{(x, y) : x^2 + y^2 \le 1\}$ the closed unit disk. Then every point on the unit circle is a global maximum, but not local one.

Theorem 9.1 *(Fermat) Suppose that $f(x, y)$ defined on an open subset U has continuous partial derivatives, and $(x_0, y_0) \in U$ is a local maximum (or a local minimum), then*

$$
\frac{\partial f(x_0, y_0)}{\partial x} = \frac{\partial f(x_0, y_0)}{\partial y} = 0. \tag{9.3}
$$

That is the gradient vector $\nabla f(x_0, y_0) = 0$.

Proof. Consider the local maximum case. There is $\varepsilon > 0$ such that $B_\varepsilon(x_0, y_0) \subset U$ and (9.1) holds. For any unit vector $\mathbf{v} = (v_1, v_2)$ and let $\gamma(t) = (x_0, y_0) + t\mathbf{v}$. Consider one variable function $g(t) = f \circ \gamma(t)$. Then $g(t) \le g(0)$ for any $t \in (-\varepsilon, \varepsilon)$ and $g'(0)$ exists by the chain rule. On the other hand

$$
g'(0) = \lim_{\substack{t \to 0 \\ t > 0}} \frac{g(t) - g(0)}{t} \le 0
$$

and

$$g'(0) = \lim_{\substack{t\to 0 \\ t<0}} \frac{g(t) - g(0)}{t} \geq 0$$

so we must have $g'(0) = 0$. While, $g'(0)$ is just the directional derivative of f in $\mathbf{v} = (v_1, v_2)$ so that

$$D_{\mathbf{v}} f(x_0, y_0) = v_1 \frac{\partial f(x_0, y_0)}{\partial x} + v_2 \frac{\partial f(x_0, y_0)}{\partial y} = 0$$

for any unit vector (v_1, v_2), which yields (9.3). ■

Any point (x_0, y_0) such that $\nabla f(x_0, y_0) = 0$ is called a *critical* (or *stationary*) point. Fermat's theorem says local extrema must be stationary points. Therefore we search for local extrema among the stationary points. Taylor's expansion allows us say more about whether a stationary point is a local extreme point or not.

To this end, we have to look at the remainder term which appears in Taylor's expansion, i.e. the term

$$f_{xx}(\xi)(x - x_0)^2 + 2f_{xy}(\xi)(x - x_0)(y - y_0) + f_{yy}(\xi)(y - y_0)^2.$$

By considering the quadratic function $a\lambda^2 + 2c\lambda + b$ whose discriminate is $4(c^2 - ab)$, we have the following

Lemma 9.2 *1) If $c^2 - ab < 0$ and $a > 0$ (so $b > 0$ as well) then*

$$a\lambda^2 + 2c\lambda\mu + b\mu^2 \geq 0$$

and equality holds if and only if $\lambda = \mu = 0$.

2) If $c^2 - ab < 0$ and $a < 0$ (so $b < 0$ as well) then $A\mathbf{v} \cdot \mathbf{v} \leq 0$

$$a\lambda^2 + 2c\lambda\mu + b\mu^2 \leq 0$$

and equality holds if and only if $\lambda = \mu = 0$.

Together with Taylor's expansion we are now in a position to derive further information about stationary points.

Theorem 9.3 *Suppose that $f(x, y)$ defined on an open subset U has continuous derivatives up to second order, and suppose $(x_0, y_0) \in U$ is a critical point: $\nabla f(x_0, y_0) = 0$.*

1) If

$$\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}\right)^2 - \frac{\partial^2 f(x_0, y_0)}{\partial x^2}\frac{\partial^2 f(x_0, y_0)}{\partial y^2} < 0, \quad \frac{\partial^2 f(x_0, y_0)}{\partial x^2} > 0 \tag{9.4}$$

then (x_0, y_0) is a local minimum.

2) If

$$\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}\right)^2 - \frac{\partial^2 f(x_0, y_0)}{\partial x^2}\frac{\partial^2 f(x_0, y_0)}{\partial y^2} < 0, \quad \frac{\partial^2 f(x_0, y_0)}{\partial x^2} < 0 \tag{9.5}$$

then (x_0, y_0) is a local maximum.

Proof. Since all partial derivatives up to second order are continuous, we can choose a small $\varepsilon > 0$ so that the open disk $B_\varepsilon(x_0, y_0) \subset U$ and (9.4) (resp. (9.5)) hold not only at $\mathbf{a} = (x_0, y_0)$ but also at any point in $B_\varepsilon(\mathbf{a})$. For any $\mathbf{x} \in B_\varepsilon(\mathbf{a})$, according to Taylor's theorem, there is $\xi \in B_\varepsilon(\mathbf{a})$ (though depending on $\mathbf{x}$) such that

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) \\ &\quad + f_{xx}(\xi)(x - x_0)^2 + 2f_{xy}(\xi)(x - x_0)(y - y_0) + f_{yy}(\xi)(y - y_0)^2 \end{aligned}$$

Suppose (9.4) holds, so it holds on $B_\varepsilon(\mathbf{a})$ for small $\varepsilon > 0$, so that

$$f_{xx}(\xi)(x - x_0)^2 + 2f_{xy}(\xi)(x - x_0)(y - y_0) + f_{yy}(\xi)(y - y_0)^2 \geq 0$$

which yields $f(\mathbf{x}) \geq f(\mathbf{a})$ on $B_\varepsilon(\mathbf{a})$ so $\mathbf{a}$ is a local minimum. ■

A natural question is, of course, what can we say if

$$\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}\right)^2 - \frac{\partial^2 f(x_0, y_0)}{\partial x^2} \frac{\partial^2 f(x_0, y_0)}{\partial y^2} \geq 0.$$

If

$$\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}\right)^2 - \frac{\partial^2 f(x_0, y_0)}{\partial x^2} \frac{\partial^2 f(x_0, y_0)}{\partial y^2} = 0 \tag{9.6}$$

then, based only on the information about the first and second partial derivatives at (x_0, y_0), we can not know the sign of

$$f_{xx}(\xi)(x - x_0)^2 + 2f_{xy}(\xi)(x - x_0)(y - y_0) + f_{yy}(\xi)(y - y_0)^2$$

appearing in the Taylor expansion, so we are in this case unable to tell whether (x_0, y_0) is a local extreme point or not.

On the other hand, if

$$\left(\frac{\partial^2 f(x_0, y_0)}{\partial x \partial y}\right)^2 - \frac{\partial^2 f(x_0, y_0)}{\partial x^2} \frac{\partial^2 f(x_0, y_0)}{\partial y^2} > 0 \tag{9.7}$$

then, by continuity, the same inequality remains to hold on a small disk near (x_0, y_0), and thus

$$f_{xx}(\xi)(x - x_0)^2 + 2f_{xy}(\xi)(x - x_0)(y - y_0) + f_{yy}(\xi)(y - y_0)^2$$

is indefinite, i.e. it can take both positive and negative values, so in this case the stationary point (x_0, y_0) is not a local extreme point, such a critical point is called a *saddle point*.

Example 9.4 *Consider $f(x, y) = \sin x + \sin y - \sin(x + y)$. Find the maximum and minimum values of f on the triangle enclosed by the x-axis, y-axis and the line $x + y = 2\pi$.*

The triangle is bounded and closed, and f is continuous, so f achieves its maximum and minimum values. The global extrema must lies on the boundary of the triangle, i.e. $x = 0$, $0 \leq y \leq 2\pi$; $y = 0$, $0 \leq x \leq 2\pi$; $x + y = 2\pi, 0 \leq x, y \leq 2\pi$, or lies in the interior of the triangle. In

this case, a global extreme point must be a local one, hence must be critical points of f. Hence we first locate the possible critical points inside the triangle by solving the following system

$$\begin{aligned}\frac{\partial f}{\partial x} &= \cos x - \cos(x+y) = 0,\\ \frac{\partial f}{\partial y} &= \cos y - \cos(x+y) = 0,\end{aligned}$$

to obtain only one critical point $(\frac{2\pi}{3},\ \frac{2\pi}{3})$ and $f(\frac{2\pi}{3},\ \frac{2\pi}{3}) = \sqrt{3}/2$. On the other hand on the boundary $f(x,y) = 0$ so $(\frac{2\pi}{3},\ \frac{2\pi}{3})$ is the global maximum.

Since

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= -\sin x + \sin(x+y),\ \frac{\partial^2 f}{\partial y^2} = -\sin y + \sin(x+y),\\ \frac{\partial^2 f}{\partial x \partial y} &= \sin(x+y),\end{aligned}$$

at $(\frac{2\pi}{3},\ \frac{2\pi}{3})$, the discriminate

$$\begin{aligned}D &= \sin^2(\frac{4\pi}{3}) - \left(-\sin\frac{2\pi}{3} + \sin\frac{4\pi}{3}\right)^2\\ &= \frac{3}{4} - 3 < 0\end{aligned}$$

and

$$\frac{\partial^2 f}{\partial x^2}(\frac{2\pi}{3}, \frac{2\pi}{3}) = -\sin\frac{2\pi}{3} + \sin\frac{4\pi}{3} = -\sqrt{3} < 0$$

so that $(\frac{2\pi}{3},\ \frac{2\pi}{3})$ is a local maximum,

There is a generalization to several variable functions. To this end we have to borrow a notion about symmetric matrices from the linear algebra. We say an $n \times n$ symmetric matrix $A = (a_{ij})$ (where $a_{ij} = a_{ji}$ for any pair (i,j)) is *positive definite* (resp. *negative definite*) if

$$A\mathbf{v} \cdot \mathbf{v} = \sum_{i,j=1}^{n} a_{ij} v_i v_j \geq 0 \quad (\text{resp.} \ \leq 0) \quad \forall \mathbf{v} = (v_1, \cdots, v_n) \in \mathbb{R}^n, \tag{9.8}$$

and equality holds if and only if $\mathbf{v} = 0$.

For a function $f(x_1, \cdots, x_n)$ of n variables with continuous partial derivatives up to second order, then the hessian matrix $D^2 f$ is an $n \times n$ matrix with entry $\frac{\partial^2 f}{\partial x_i \partial x_j}$ at the ith row and jth column, i.e.

$$D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix} \tag{9.9}$$

which a symmetric matrix-valued function.

Theorem 9.5 *Suppose $f(\mathbf{x})$ is a function with n variables $\mathbf{x} = (x_1, \cdots, x_n)$ defined on an open subset $U \subset \mathbb{R}^n$ which has continuous partial derivatives up to second order. Let $\mathbf{a} = (a_1, \cdots, a_n)$ be a critical point: $\nabla f(\mathbf{a}) = 0$.*

1) If the hessian matrix $D^2 f(\mathbf{a})$ is positive definite, then $\mathbf{a}$ is a local minimum of f.

2) If the hessian matrix $D^2 f(\mathbf{a})$ is negative definite, then $\mathbf{a}$ is a local maximum of f.

The proof follows from a discussion via Taylor's expansion at the critical point $\mathbf{a}$.

10 Lagrange's multipliers

In this part we develop a method of locating relative local extrema. Let us first consider the question with three variables, and consider the following problem. Let $f(x, y, z)$ be a function defined on a subset $U \subset \mathbb{R}^3$. We wish to locate the local extrema of $f(x, y, z)$ subject to the following constraint

$$F(x, y, z) = 0. \tag{10.1}$$

We say $(x_0, y_0, z_0) \in U$ is a (relative) local minimum subject to (10.1) if $F(x_0, y_0, z_0) = 0$ and there is a small ball B_ε centered at (x_0, y_0, z_0) with radius $\varepsilon > 0$ such that $f(x, y, z) \geq f(x_0, y_0, z_0)$ for every $(x, y, z) \in B_\varepsilon$ which satisfies (10.1).

Theorem 10.1 *Let $f(x, y, z)$ and $F(x, y, z)$ be two functions on an open subset $U \subset \mathbb{R}^3$. Suppose that both functions f and F have continuous partial derivatives, and the gradient vector field $\nabla F \neq 0$ on U. Let $(x_0, y_0, z_0) \in U$ be a local maximum or local minimum of $f(x, y, z)$ subject to the constraint (10.1). Then there is a real number λ such that $\nabla f(x_0, y_0, z_0) = \lambda \nabla F(x_0, y_0, z_0)$.*

Proof. Since $\nabla F \neq 0$ on U, the equation (10.1) defines a surface

$$S = \{(x, y, z) \in U : F(x, y, z) = 0\}.$$

By assumptions, $(x_0, y_0, z_0) \in S$ is a local maximum or minimum of the restriction of the function f over S. Given any differential curve $\gamma(t) = (x(t), y(t), z(t))$ lying on the surface S, passing through (x_0, y_0, z_0), i.e.

$$F \circ \gamma(t) = 0 \quad \forall t \in (-\varepsilon, \varepsilon), \ \gamma(0) = (x_0, y_0, z_0),$$

and consider $h(t) = f \circ \gamma(t)$. Then by the definition of relative local extrema, 0 is a local maximum or minimum of the function $h(t)$. Therefore, by Fermat's theorem, $h'(0) = 0$. On the other hand, according to the chain rule

$$h'(0) = \nabla f(\gamma(0)) \cdot \gamma'(0) = 0,$$

which means that $\nabla f(x_0, y_0, z_0)$ is perpendicular to $\gamma'(0)$. Since $\gamma(t)$ is any curve lying on the surface S passing through (x_0, y_0, z_0), so that $\gamma'(0)$ can be any tangent vector to S at (x_0, y_0, z_0). Therefore $\nabla f(x_0, y_0, z_0)$ must be perpendicular to the tangent plane of S at (x_0, y_0, z_0). It follows that $\nabla f(x_0, y_0, z_0)$ either equals 0 or $\nabla f(x_0, y_0, z_0) \neq 0$ is normal to S at (x_0, y_0, z_0). On the other hand a normal vector to S at (x_0, y_0, z_0) is $\nabla F(x_0, y_0, z_0)$, therefore $\nabla f(x_0, y_0, z_0)$ and $\nabla F(x_0, y_0, z_0)$ are parallel. Since $\nabla F(x_0, y_0, z_0) \neq 0$, so there is λ such that $\nabla f(x_0, y_0, z_0) = \lambda \nabla F(x_0, y_0, z_0)$. ■

As a by-product, we have proved that if $(x_0, y_0, z_0) \in S$ is a relative local maximum or minimum of $f(x, y, z)$ along S (i.e. satisfying the constraint (8.5)), then $\nabla f(x_0, y_0, z_0)$ is perpendicular to the level surface $S : F(x, y, z) = 0$.

According to the previous theorem, in order to find the constrained extrema of f we should look among those $(x, y, z) \in U$ and real number λ which satisfy the following system

$$\begin{cases} \nabla f(x, y, z) = \lambda \nabla F(x, y, z), \\ F(x, y, z) = 0. \end{cases} \tag{10.2}$$

[We often assume that $\nabla F(x, y, z) \neq 0$]. Of course we are interested in those $(x, y, z) \in U$ such that there is a real number λ which solve the system (10.2). In practice, we need to solve (x, y, z), but there is no need to know the explicit value λ. The constant λ introduced here to help us to locate the relative extrema is called a *Lagrange multiplier*.

Introduce a function $G(x, y, z, \lambda) = f(x, y, z) - \lambda F(x, y, z)$. Then the system (10.2) may be written as

$$\frac{\partial G}{\partial x} = \frac{\partial G}{\partial y} = \frac{\partial G}{\partial z} = \frac{\partial G}{\partial \lambda} = 0$$

which means a solution (x, y, z, λ) to (10.2) is just a critical point of $G(x, y, z, \lambda)$.

Example 10.2 *Maximize $f(x, y, z) = x + y$ subject to the constraint $x^2 + y^2 + z^2 = 1$.*

To use the method of Lagrange multipliers, set

$$G(x, y, z, \lambda) = x + y - \lambda \left(x^2 + y^2 + z^2 - 1\right)$$

and look for the critical points of G by solving the system

$$\begin{aligned} \frac{\partial G}{\partial x} &= 1 - 2\lambda x = 0, \\ \frac{\partial G}{\partial y} &= 1 - 2\lambda y = 0, \\ \frac{\partial G}{\partial z} &= -2\lambda z = 0, \\ \frac{\partial G}{\partial \lambda} &= x^2 + y^2 + z^2 - 1 = 0. \end{aligned}$$

The first equation implies that $\lambda \neq 0$, so from the first and second equations, we obtain $z = 0$, $x = y = \frac{1}{2\lambda}$, substituting them to the constraint to obtain

$$\left(\frac{1}{2\lambda}\right)^2 + \left(\frac{1}{2\lambda}\right)^2 + 0^2 = 1$$

so that $\frac{1}{2\lambda} = \pm\sqrt{\frac{1}{2}}$. Thus there are two possible relative extrema $(\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}, 0)$ and $(-\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{2}}, 0)$. Since the sphere $S : x^2 + y^2 + z^2 = 1$ is compact (bounded and closed), and the function $f(x, y, z) = x + y$ is continuous, so it must achieve it maximum and minimum values [We will prove this kind of statements in Prelims Analysis II]. Therefore the maximum of f subject to the constraint is

$$f(\sqrt{\frac{1}{2}}, \sqrt{\frac{1}{2}}, 0) = \sqrt{2}$$

while

$$f(-\sqrt{\frac{1}{2}},-\sqrt{\frac{1}{2}},0)=-\sqrt{2}$$

is the constrained minimum values of f over the unit sphere.

To conclude our discussion, let us describe the general form of the Lagrange multipliers. Suppose that $f(x_1,\cdots,x_n)$ and $F_1(x_1,\cdots,x_n)$, $\cdots$, $F_k(x_1,\cdots,x_n)$ are functions with n variables defined on an open subset $U\subset\mathbb{R}^n$, where $n,k\in\mathbb{N}$. Suppose that f, $F_1,\cdots,F_k$ have continuous partial derivatives. Then the local extrema of $f(x_1,\cdots,x_n)$ subject to the following constraints:

$$\begin{cases} F_1(x_1,\cdots,x_n)=0, \\ \qquad\cdots \\ F_k(x_1,\cdots,x_n)=0 \end{cases}$$

are solutions to the following system

$$\frac{\partial G}{\partial x_1}=\cdots=\frac{\partial G}{\partial x_n}=\frac{\partial G}{\partial \lambda_1}=\cdots=\frac{\partial G}{\partial \lambda_k} \tag{10.3}$$

where

$$\begin{aligned} G(x_1,\cdots,x_n,\lambda_1,\cdots,\lambda_k) &= f(x_1,\cdots,x_n)-\lambda_1F_1(x_1,\cdots,x_n) \\ &\quad -\cdots-\lambda_kF_k(x_1,\cdots,x_n), \end{aligned}$$

the additional constants $\lambda_1,\cdots,\lambda_k$ are called the Lagrange multipliers.

Example 10.3 *Find the extreme points of $f(x,y,z)=x+y+z$ subject to the conditions $x^2+y^2=2$ and $y^2+z^2=2$.*

Construct function

$$G(x,y,z,\lambda_1,\lambda_2)=x+y+z-\lambda_1(x^2+y^2-2)-\lambda_2(y^2+z^2-2).$$

We want to solve, in order to locate extreme points, the following system

$$\begin{aligned} \frac{\partial G}{\partial x} &= 1-2\lambda_1x=0, \\ \frac{\partial G}{\partial y} &= 1-2(\lambda_1+\lambda_2)y=0, \\ \frac{\partial G}{\partial y} &= 1-2\lambda_2z=0, \end{aligned}$$

together with the constraints $x^2+y^2=2$ and $y^2+z^2=2$. $\lambda_1,\lambda_1,\lambda_1+\lambda_2\neq 0$ and $x=\frac{1}{2\lambda_1}$, $z=\frac{1}{2\lambda_2}$ and $y=\frac{1}{2(\lambda_1+\lambda_2)}$. From the constraints we deduce that $x=\pm z$ which implies that $\lambda_1=\lambda_2$ as $\lambda_1+\lambda_2\neq 0$. Hence $x=z=\frac{1}{2\lambda_1}$ and $y=\frac{1}{4\lambda_1}$. Using again the constraints to obtain that

$$\left(\frac{1}{2\lambda_1}\right)^2+\left(\frac{1}{4\lambda_1}\right)^2=2$$

which leads to the solutions $\frac{1}{\lambda_1}=\pm 4\sqrt{\frac{2}{5}}$. Thus possible constrained extreme points are

$$\left(2\sqrt{\frac{2}{5}},\sqrt{\frac{2}{5}},2\sqrt{\frac{2}{5}}\right) \text{ and } \left(-2\sqrt{\frac{2}{5}},-\sqrt{\frac{2}{5}},-2\sqrt{\frac{2}{5}}\right)$$

at which the function f achieves the relative maximum and minimum values respectively.